Research Article
Towards Automatic Image Exposure Level Assessment

Lin Zhang,1 Xilin Yang,1 Lijun Zhang,2 Xiao Liu,2 Shengjie Zhao,1 and Yong Ma3

1 School of Software Engineering, Tongji University, Shanghai, China
2 College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst 01003, MA, USA
3 School of Computer Information Engineering, Jiangxi Normal University, Nanchang, China

Correspondence should be addressed to Lin Zhang; cslinzhang@tongji.edu.cn and Lijun Zhang; lijunzhang@cs.umass.edu

Received 18 July 2020; Revised 8 October 2020; Accepted 10 November 2020; Published 23 November 2020

Academic Editor: Nazrul Islam

Copyright © 2020 Lin Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The quality of acquired images can be surely reduced by improper exposures. Thus, in many vision-related industries, such as imaging sensor manufacturing and video surveillance, an approach that can routinely and accurately evaluate exposure levels of images is in urgent need. Taking an image as input, such a method is expected to output a scalar value, which can represent the overall perceptual exposure level of the examined image, ranging from extremely underexposed to extremely overexposed. However, studies focusing on image exposure level assessment (IELA) are quite sporadic. It should be noted that blind NR-IQA (no-reference image quality assessment) algorithms or metrics used to measure the quality of contrast-distorted images cannot be used for IELA. The root reason is that though these algorithms can quantify quality distortion of images, they do not know whether the distortion is due to underexposure or overexposure. This paper aims to resolve the issue of IELA to some extent and contributes to two aspects. Firstly, an Image Exposure Database (IEpsD) is constructed to facilitate the study of IELA. IEpsD comprises 24,500 images with various exposure levels, and for each image a subjective exposure score is provided, which represents its perceptual exposure level. Secondly, as IELA can be naturally formulated as a regression problem, we thoroughly evaluate the performance of modern deep CNN architectures for solving this specific task. Our evaluation results can serve as a baseline when the other researchers develop even more sophisticated IELA approaches. To facilitate the other researchers to reproduce our results, we have released the dataset and the relevant source code at https://cslinzhang.github.io/imgExpo/.

1. Introduction

Exposure is the total amount of light falling on a photographic medium when capturing an image [1]. Improper exposure will inevitably reduce the quality of the acquired images, e.g., bringing contrast reduction. Thus, how to assess exposure levels of images (videos) and to correct ill-exposed images (videos) are of paramount importance in the research area of multimedia.

An exposure distortion is understood as the overall quality degradation caused by improper exposure. In many industrial fields, a method that can accurately assess the exposure levels of images is in urgent need [2–5]. For example, almost all the modern digital cameras can work in “autoexposure” mode [2]. When the user is taking images with this mode, the camera will automatically adjust relevant hardware parameters (such as the aperture, the shutter speed, and the electronic gain [6]) using a particular autoexposure algorithm to make the collected images have proper exposure levels. Obviously, in order to verify the performance of an autoexposure algorithm, a method that can accurately assess the exposure levels of acquired images is indispensable. Another example commonly encountered is in video surveillance. For video surveillance, it is very common that lighting conditions are out of the adaptive capacity of the camera. Hence, it is quite necessary to continuously monitor the exposure level of the acquired video to determine its quality [4].

At present, commonly adopted approaches of judging whether an image is properly exposed are based on the experience of the photographers. These kinds of schemes are of course costly and inefficient, lack robustness, and cannot be applied to systems requiring real-time
exposure level scores. Hence, there is an urgent need to
develop computational image exposure metrics.
This work tries to solve the problem of IELA (Image Exposure Level Assessment) to some extent. The ultimate goal is to obtain a computerized model that can objectively and effectively predict the overall exposure level of any given image, and the prediction results are anticipated to correlate well with human subjective judgements. The target algorithm should quantify exposure in a meaningful manner, which means that the same predicted exposure score should preferably correspond to the same exposure level across different image contents. Such an IELA algorithm has many potential applications. For example, it could be explored to measure or to optimize the performance of autoexposure models, which are of paramount importance for imaging sensor manufacturing industries.

In order to more clearly demonstrate the objectives of our work, in Figure 1, we present six images and give their exposure scores predicted by our proposed approach IEM$^{\text{SN}}$ (short for “Image Exposure Metric with ShuffleNet”; refer to Section 4 for details). It should be pointed out that exposure scores predicted by IEM$^{\text{SN}}$ can vary continuously from $-1$ to $+1$. $-1$ implies that the assessed image is extremely underexposed, $0$ implies that it is correctly exposed, and $+1$ implies that it is extremely overexposed. The more the exposure score deviates from $0$, the more serious the exposure distortion is. Using IEM$^{\text{SN}}$, the predicted exposure scores of Figures 1(a)–1(f) are $-0.8870$, $-0.5043$, $-0.2577$, $0.1368$, $0.4739$, and $0.5697$, respectively. This example demonstrates that IEM$^{\text{SN}}$’s prediction results of images’ exposure levels correlate consistently with human judgements.

The rest of this article is organized as follows. Section 2 introduces the related work, our motivations, and our contributions. Section 3 presents details of IE$p$D (short for “Image Exposure Database”), which is our newly established benchmark dataset for the study of IELA. Section 4 presents our DCNN-based image exposure level assessment model, IEM$_X$. Experimental results and related discussions are presented in Section 5. Finally, conclusions are provided in Section 6.

2. Related Work and Our Contributions

In this section, we first review some representative studies most relevant to our work, including existing approaches for IELA, approaches for no-reference (NR) quality assessment of contrast-distorted images, and approaches for blind NR image quality assessment (NR-IQA). Then, our motivations and contributions are presented.

2.1. Existing Approaches for IELA. At present, the work that specializes in IELA is quite sporadic. Human experience suggests that an image’s exposure level could be characterized by its luminance histogram. It is generally believed that the histogram of a correctly exposed image spreads over the whole range of luminance; by contrast, histograms of overexposed (underexposed) images are shifted to the bright (dark) sides. Moreover, the higher the exposure distortion is, the more significant will be the shift. Several IELA approaches were proposed in the literature just based on this hypothesis. In Liu et al.’s patent [7], three quantities “center,” “centroid,” and “effective width” are first extracted from the image’s luminance histogram and then the exposure level is derived from them using predefined rules. Based on the similar idea as Liu et al.’s invention, Rychagov and Efimov [8] patented a method for exposure estimation by comparing the mean of the illuminance histogram with predefined thresholds. In Romania et al.’s approach [4], the average luminance of three blocks with the highest mean luminance is regarded as the luminance upper bound $L_U$ and the average luminance of three blocks with the lowest mean luminance is regarded as the luminance lower bound $L_L$. Then, the exposure metric is calculated as $-(L_U + L_L)/2$.

2.2. Approaches for NR Quality Assessment of Contrast-Distorted Images. In most cases, improper exposure can reduce the contrast of the acquired images. Hence, studies focusing on NR quality assessment of contrast-distorted images are quite relevant to our work. The recent progress made in this area is briefly reviewed here.

On seeing that a database specially dedicated to contrast-distortion assessment was lacking, Gu et al. [10] established a database comprising contrast-changed images and their associated subjective ratings.

With respect to quality assessment models of contrast-distorted images, existing schemes can be roughly classified into two categories: the ones based on supervised learning (SL) and the ones not based on SL. Representative approaches based on SL include [11–14]. In [12], Fang et al. first derived five NSS models (in the form of probability density functions) based on the moment (mean, standard deviation, skewness, and kurtosis) and entropy features from images in SUN2012 [15]. Then, for any given image, a set of five likelihood features can be extracted based on learned NSS models. Finally, they adopted SVR (support vector regression) to find the mapping between the feature vectors and the perceptual quality scores. Inspired by Fang et al.’s idea [12], both Ahmed et al.’s work [11] and Wu et al.’s work [13] followed the similar “features + SVR” framework. In [11], Ahmed and Der extended the 5-D feature vector proposed in [12] to a 6-D one by introducing a new directional contrast feature derived from the curvelet domain. In [13], for feature extraction, Wu et al. extracted a 7-D feature vector (the image mean, the image variance, the image skewness, the image kurtosis, the image entropy, the mean of the phase congruency map [16], and the entropy of the phase congruency map) from each image. In Xu and Wang’s approach [14], a 4-D feature vector, consisting of the perceptual contrast of the image, the skewness, the variance, and the intensity distribution number, is extracted from each image. Concerning the regression model mapping the feature vectors to perceptual quality scores, they resorted to a three-layer BP neural network.

Panetta et al.’s approach [17] and Gu et al.’s approach [18] are two eminent schemes for quality assessment of
contrast-distorted images, which are not based on SL. In Panetta et al.’s approach [17], the image is first partitioned into blocks. Then, a local quality measure is derived for each block from its maximum and minimum luminance values. Finally, an overall single measure is obtained from local measures based on the PLIP (parameterized logarithmic image processing) model [19]. In [18], Gu et al. first removed predictable regions from the image and then they regarded the entropy of regions with maximum information as the local quality measure. They also derived a global quality measure by comparing the image’s histogram with the uniformly distributed histogram of maximum information. Finally, an overall quality score was generated as the weighted mean of local and global measures.

2.3. Approaches for Blind NR-IQA. Another research area related to our work is blind NR-IQA, which aims to devise algorithms to predict the image’s perceptual quality without knowing its high-quality reference nor its quality distortion type. Hence, the recent progress made in this area will be reviewed as well.

There exist several publicly approved databases in the area of NR-IQA. Among them, LIVE [20], CSIQ [21], LIVE Multiply Distorted (MD) [22], TID2013 [23], CID2013 [24], and LIVE Challenge [25] are most commonly used in recent studies.

With respect to blind NR-IQA models, most of the existing ones are “opinion aware,” meaning that they are obtained by training on a dataset comprising quality-distorted images and the corresponding subjective scores. Typical approaches belonging to this category include [26–33], and they have similar architectures. At the training phase, the set of feature vectors is first extracted from the training images, and then a regression model that maps the feature vectors to the associated subjective scores is learned. At the testing phase, given an image \( f \) to be assessed, its feature vector is extracted first and then is input into the learned regression model at the training phase to predict \( f \)'s objective quality score. Different kinds of regression models are adopted in these methods, including the SVR [28, 31–33], the BP (backpropagation) neural networks [27], and the deep neural networks [26, 29, 30, 34].

Having noticed the disadvantages of opinion-aware blind NR-IQA models with respect to the generalization ability and training sample collection, some researchers proposed adding new vectorized labels to aid evaluation [35], and some researchers began to develop opinion-unaware IQA models. These kinds of models do not rely on quality-distorted training images nor subjective scores. Some eminent studies in this research direction have been reported. In [36], Mittal et al. proposed the Natural Image Quality Evaluator (NIQE) model. Given an image \( f \) to be evaluated, NIQE first extracts from it a set of local features and then fits them to a multivariate Gaussian (MVG) model. The perceptual quality of \( f \) is expressed as the distance between its MVG model and the MVG model learned from the image set composed of high-quality natural images. Inspired by [36], Zhang et al. [37] introduced three additional types of quality-aware features. At the test stage, on each patch of a test image, a best-fit MVG model is computed online. The overall quality score of the test image is then obtained through pooling the patch scores by averaging them. In [38], Xue et al. synthesized a virtual image set, in which the perceptual scores of the quality-distorted images were provided by FSIM (a full-reference IQA algorithm) [39]. Then, an NR-IQA model was learned from the established dataset by patch-based clustering. In [40], Wu
et al. first extracted local features using LBP (Local Binary Pattern) [41] and then obtained statistics from local image patterns, which are believed to have the capability to discriminate high-quality natural images from distorted ones. In Wu et al.’s approach [42], a feature fusion scheme is first introduced by combining the image’s statistical information from multiple domains and color channels. Then, the predicted image quality is generated by label transfer (LT), where a query image’s KNNs (k-nearest neighbors) are searched for from some annotated images. Gu et al.’s approach [43] is based on vector regression and an object-oriented pooling strategy. By extending LBP, Freitas et al. designed the OCPP (Orthogonal Color Planes Pattern) descriptor, and they used the statistics of the OCPP descriptor to characterize image quality [44].

2.4. Our Motivations and Contributions. Through the literature survey, it can be found that though IELA is a problem of paramount importance, systematic and in-depth studies in this field are still lacking.

First, all the existing IELA metrics [4, 7–9] are derived from luminance histograms, and accordingly, their shared drawback is that they are not image content-independent. In most cases, a useful IELA metric is expected to be content-independent. However, existing IELA metrics do not satisfy this requirement because they are totally defined on luminance histograms. As shown in Figures 2(a)–2(c), three images have the same image content, but their histograms have different distribution patterns because of their different exposure levels. The histogram of the properly exposed image (Figure 2(a)) expands over the whole luminance range, while the histogram of the overexposed (underexposed) image moves to the right (left) as shown in Figure 2(b) (Figure 2(c)). Existing IELA methods [4, 7–9] were designed precisely based on the assumption that images’ perceptual exposure levels could be well characterized by their luminance histograms. However, this assumption becomes problematic when applied to images taken from various scenes. As shown in Figures 2(d)–2(f), though all three images are exposed correctly, their histogram distribution patterns differ apparently from each other owing to their different contents. As a consequence, when dealing with images similar to Figures 2(d)–2(f), IELA metrics totally based on luminance histograms [4, 7–9] would yield erroneous prediction results. In a word, the outputs of [4, 7–9] depend on image contents, and consequently, their accuracy in measuring the image exposure level is quite limited.

Second, blind NR-IQA algorithms or metrics used to measure the quality of contrast-distorted images cannot be used for IELA. When an image with improper exposure is fed into these algorithms, they can quantify its quality degradation caused by improper exposure, but the evaluation results cannot indicate whether the degradation is due to underexposure or overexposure. This fact is further illustrated by examples shown in Figure 3. By perceptual evaluation, it can be found that the images in Figures 3(a)–3(c) are underexposed, properly exposed, and overexposed, respectively. Their objective scores evaluated by “NIQMC” [10], “CS-BIQA” [33], and “IEM\textsubscript{SN}” are presented in Table 1. NIQMC is a state-of-the-art metric to measure the quality of contrast-distorted images, and a higher NIQMC score indicates higher contrast. CS-BIQA is a representative modern blind NR-IQA model, and a lower CS-BIQA score indicates higher quality. IEM\textsubscript{SN} is our proposed IELA model (refer to Section 4 for details) trained on our established dataset used for the IELA study (refer to Section 3 for details). From Table 1, it can be seen that NIQMC and CS-BIQA can characterize an image’s quality degradation quite well. However, whether the examined image is underexposed or overexposed cannot be reflected from their results. By contrast, the proposed IELA model IEM\textsubscript{SN} can accurately and unambiguously evaluate the exposure levels of given images. The interpretation of IEM\textsubscript{SN}’s output can be found in Section 1.

Third, there is no publicly available benchmark dataset specially designed to study the IELA problem. To design and evaluate IELA approaches, such a dataset is actually indispensable.

This work attempts to fill the aforementioned research gaps partially. The major contributions are briefed as follows.

(1) To facilitate training and testing IELA models, a benchmark dataset, namely, IEpsD (Image Exposure Database), has been established. IEpsD contains 24,500 images with different exposure levels. 3,500 of them were collected from the real-world while the other 21,000 ones were synthesized from properly exposed source images by using our exposure simulation pipeline. For each image in IEpsD, a corresponding subjective score is provided to represent its perceptual exposure level. To our knowledge, IEpsD is the first large-scale benchmark dataset established for the study of IELA. In our experiments, synthetic images in IEpsD are used for training IELA models, while real-world ones of IEpsD are used for testing. For more details about IEpsD, refer to Section 3.

(2) The problem of IELA can be formulated as a regression problem from the input image to its subjective exposure score, which can be naturally solved by DCNNs (Deep Convolutional Neural Networks [45]). Hence, in this paper, a DCNN-based model IEM\textsubscript{X} (Image Exposure Metric using X) is proposed for IELA, which can learn an end-to-end mapping from images to their subjective exposure scores. Here “X” denotes a concrete DCNN architecture used. In experiments, a thorough evaluation has been conducted to assess the performance of modern DCNN architectures for IELA in the framework of IEM\textsubscript{X} (refer to Section 5 for details).

We have released IEpsD and the relevant source code at https://cslinzhang.github.io/imgExpo/ to facilitate the other researchers to reproduce our results.

A preliminary version of this manuscript has been presented on ICME 2018 [46]. The following improvements are made in this version: (1) the database IEpsD is
substantially extended and a more reasonable way to perform the subjective evaluation of exposure levels is adopted; (2) the performance of blind NR-IQA models and metrics used to measure the quality of contrast-distorted images for addressing the problem of IELA is thoroughly investigated and analyzed; (3) thorough performance evaluation of modern DCNN architectures in the framework of IEMX is conducted; and (4) more competing IELA models are evaluated in experiments.

3. IEpsD: A Benchmark Dataset for IELA

As stated in Section 2, in view of the fact that a database specially dedicated to IELA still lacks in the community, we are motivated to establish such a dataset in this work. This section will discuss details about the establishment of our image exposure dataset IEpsD and its practical use. By collecting and synthesizing images of various exposure levels from different shooting scenes, IEpsD finally contains 24,500 images. Additionally, for each image in IEpsD, we provide it with a subjective score which is expected to represent its perceptual exposure level.

Three phases were involved in constructing IEpsD, including collection of real-world images, generation of synthetic images, and finally subjective evaluation.

3.1. Collection of Real-World Images. In order to accurately quantify an IELA algorithm’s prediction accuracy on real data, IEpsD should include a large number of real-world
images. When taking these images, the shooting scenes need to be as diverse as possible, meaning that they should cover different kinds of objects (humans, plants, animals, human-made objects, etc.), different periods of the day (morning, noon, afternoon, evening, and night), different lighting conditions, and different shooting distances. Taking these factors into consideration, we finally collected images from 500 shooting scenarios which were carefully planned. An iPhone7 Plus mobile phone was used for image collection.

For digital cameras, exposure levels can be modulated in three ways. The first way is by enlarging or shrinking the aperture. The larger the iris aperture is, the more light reaches the imaging sensor in a fixed period of time. The second way is by adjusting the ISO sensitivity. The last way is by varying the exposure time. To simplify data collection operations, we only changed the exposure time and kept the other factors unchanged to obtain 7 different exposure results, ranging from extremely underexposed to extremely overexposed.

In the end, 3,500 (7 × 500) real-world images were collected, and we denote the dataset formed by them by IEpsD_R. Thumbnails of 28 sample images selected from IEpsD_R are shown in Figure 4. In Figure 4, from top to bottom, images in each row belong to one specific shooting scenario; from left to right, the exposure levels are changing from “extremely overexposed” to “extremely underexposed.”

3.2. Generation of Synthetic Images with Various Exposure Levels. To get an IELA model with a satisfying generalization capability, a large-scale dataset, comprising a large number of images with various exposure levels, is indispensable for training. Unfortunately, establishing such a real-world dataset is extremely costly and laborious. In order to resolve this contradiction, we propose to use synthetic images for training IELA models. Actually, in the community of computer vision, researchers have recently found that the use of synthetic images can effectively alleviate the problem of insufficient real training data. This has spurred the development of pipelines for synthesizing photo-realistic images. Synthetic data have already been explored to train models to tackle the problems such as object detection [47], semantic segmentation [48], optical flow estimation [49], and so on. In this paper, we propose a novel method for generating synthetic images with various exposure levels from properly exposed source images.

Suppose that I is a given properly exposed source image. A synthetic image $\tilde{I}$ with a different exposure level could be created by modulating I’s illumination and saturation channels. In order to manipulate the illumination and saturation channels separately, we first convert I from the RGB space to the HSV space. Denote the illumination channel and the saturation channel of I by $I_i$ and $I_s$, respectively. Similarly, denote the illumination channel and the saturation channel of $\tilde{I}$ by $\tilde{I}_i$ and $\tilde{I}_s$, respectively. $\tilde{I}_i$ is generated by adjusting $I_i$ as

$$\tilde{I}_i(x) = I_i(x) + \theta,$$

where x denotes the spatial location and $\theta$ is a global parameter controlling the amount of illumination adjustment. $\theta$ should be positive when simulating an overexposed image, while it should be negative when simulating an underexposed one.

In addition, $I_s$ needs to be adjusted to $\tilde{I}_s$ accordingly. As suggested by Romaniak et al. [4], the mapping function between I(x)’s exposure level $E_{in}(x)$ and its saturation value $I_s(x)$ conforms to an inverse asymmetric logit function (IALF) given by the following equation:

$$E_{in}(x) = \frac{\ln((I_s(x)^{1/\epsilon})/1-I_s(x)^{1/\epsilon}) - b}{a}, \quad (2)$$

where $a$, $b$, and $c$ are three given constants. $\tilde{I}(x)$’s exposure level $E_{out}(x)$ can be obtained by shifting $E_{in}(x)$ by a desired offset eps, i.e.,

$$E_{out}(x) = E_{in}(x) + \text{eps}. \quad (3)$$

At last, $\tilde{I}(x)$’s saturation value $\tilde{I}_s(x)$ can be calculated by the following asymmetric logit function (ALF):

$$\tilde{I}_s(x) = \frac{e^{a(E_{out}(x)+\text{eps})}}{1 + e^{a(E_{out}(x)+\text{eps})}}^c. \quad (4)$$

Putting equations (2)–(4) together, we can get the formula for adjusting $I_i$ to $\tilde{I}_i$ as

$$\tilde{I}_i(x) = \frac{e^{a(E_{out}(x)+\text{eps})}}{1 + e^{a(E_{out}(x)+\text{eps})}}^c + \text{eps} \quad (5)$$

In our implementation, $a$ is set to $-3.2$ and $c$ is set to $0.4$. By altering the values of parameters $\theta$ and eps, we can synthesize a series of $\tilde{I}$’s variants with different exposure levels. Specifically, to construct IEpsD, seven exposure levels were synthesized. Alternatively, in other words, from each properly exposed source image, seven images (including the source image itself) having different exposure levels, ranging from “extremely underexposed” to “extremely overexposed,” were synthesized. Sample synthetic images generated by our proposed scheme are shown in Figure 5. In Figure 5, images in the first column are the properly exposed source images, based on which the synthetic ones are generated. Columns 2–4 are the synthetic results of overexposed images while columns 5–7 show the synthetic results of underexposed ones. By visual inspection, it can be found that using our proposed scheme, the appearance of synthetic images looks quite natural and correlates well with human perception.

To establish the synthetic image dataset, we collected a set of properly exposed images from the Internet. Four volunteers (postgraduate students from Tongji University, Shanghai, China) were involved, and each of them was asked to search for 1000 high-quality images covering four categories: people, plants, animals, and man-made objects. Then, each of the 4000 collected images was visually examined by seven volunteer observers (undergraduate students from Tongji University). If no fewer than five of the seven observers confirmed that the image being examined was properly exposed, then the image was retained. Through this way, 3000 images were thus selected, and they were used as
source images for generating the synthetic ones. Note that none of the images used here is included in IEpsD_R. Finally, using our proposed synthetic image generation model, 21,000 ($7 \times 3000$) synthetic images were generated, and we denote the dataset formed by them by IEpsD_S.

To demonstrate the reliability of the synthetic images in the established dataset, the comparison between the real images and the generated images in terms of the brightness, which could reflect the applicability of simulated exposure levels to some extent, is conducted.

Figure 6 shows the comparison of the average brightness distributions between the real images and the synthetic images. The $X$-coordinate is the normalized average brightness value of the image. The $Y$-coordinate is the number of images. The seven colors of the bar represent different exposure levels. It can be found from the figure that the distributions of the average brightness of the real images and the synthetic images under various exposure levels are similar, which indicates that the algorithm proposed in this paper for generating synthetic images is reasonable and effective.

The final dataset IEpsD comprises two parts, IEpsD_R (formed by real-world images) and IEpsD_S (formed by synthetic images).

3.3. Subjective Evaluation for IEpsD. When the image set IEpsD is ready, the next step is to assign a subjective score to each image in IEpsD, which can reflect its perceptual exposure level.
The subjective evaluations were conducted following a single-stimulus strategy [50]. The reason for choosing a single-stimulus methodology instead of a double-stimulus one was that the number of images to be assessed was extremely huge for a double-stimulus study (we evaluated a total of 24,500 images). Subjective evaluations were performed on identical workstations. Monitors of workstations were all 22-inch LCD monitors, and their screen resolutions were all set to 1920 × 1080. Evaluations were conducted in an indoor environment with normal illuminations. Matlab software was developed to assist the subjective study. The lab setup is illustrated in Figure 7. Subjects taking part in the subjective evaluation were all undergraduate students of Tongji University, and they were inexperienced with image exposure level assessment. The number of subjects evaluating each image was 20.

For each participant, we explained to him/her the goal of the experiment and also the experimental procedure. We also showed each participant the approximate range of image exposure levels and the corresponding scoring results in a short training session. It needs to be noted that we used different images in the training session from those used in the actual experiment. During the subjective evaluation, images were displayed to a subject in random order, and for different subjects, the randomization processes were different. A subject reported his/her judgement of the exposure level by dragging a slider on a quality scale. The quality scale was marked both numerically and textually and was divided into five equal portions, which were labeled as “Extremely Underexposed,” “Underexposed,” “Normally Exposed,” “Overexposed,” and “Extremely Overexposed,” respectively. After the subject evaluated the image, by uniformly mapping the entire quality scale to the range [−50, 50], the position of the slider was converted into an integer exposure score. By this way, raw exposure scores obtained from subjects were integers falling in the range [−50, 50]. The closer the score is to “0,” the more likely the image is exposed normally. A score smaller than “0” means the examined image is underexposed, and a score above “0” means the image is overexposed. Moreover, the more the exposure score deviates from “0,” the more serious the exposure distortion is.

Next, some postprocessing steps were applied to subjects’ raw scores. At first, to eliminate the influence of different subjective evaluation standards of subjects, the raw scores $d_{ij}$ were normalized as

$$z_{ij} = \frac{d_{ij} - \bar{d}_i}{\sigma_i},$$  

where $d_{ij}$ is the exposure score of the image $I_j$ given by the $i$th subject, $\bar{d}_i$ is the mean score of the $i$th subject, $\sigma_i$ is the
standard deviation of his/her scores for all images, and $z_{ij}$ is the normalized score of the image $I_j$ given by the $i$th subject. Then, we used a strategy similar to the one mentioned in [51] to filter out those heavily biased subjective scores, which satisfied

$$z_{ij} - \overline{z_j} > T \cdot \sigma_j,$$  \hspace{1cm} (7)

where $\overline{z_j}$ is the mean of the normalized scores of $I_j$, $T$ is a threshold constant, and $\sigma_j$ is the standard deviation of $I_j$’s normalized scores. The mean of the remaining evaluation scores of $I_j$ was deemed as $I_j$’s subjective exposure score $s_j$:

$$s_j = \frac{1}{N_j} \sum_{i=1}^{N_j} z_{ij},$$ \hspace{1cm} (8)

where $N_j$ is the number of valid subjective scores for $I_j$. Finally, $s_j$ is linearly rescaled to the range $[-1, 1]$.

Now, for each image $I_j$ in IEpsD, we get a subjective score $s_j$ that reflects its perceptual exposure level.

### 3.4. Practical Use of IEpsD

In addition to being used for IELA research, our dataset also has great potential in lots of relevant fields like high dynamic range (HDR) and image exposure correction.

HDR images can provide more dynamic range and image details and reflect the visual effects better in the real environment than ordinary images. The most common way to capture HDR images is to take a series of low dynamic range (LDR) images at different exposures and then merge them into an HDR image [52]. IEpsD contains sequences of images which are very diverse and often contain complex scenes with multiple objects. Such images usually possess the same content while having different exposure levels, so they can be used to generate HDR images and to conduct related studies.

For image exposure correction, IEpsD can be used as a benchmark dataset to evaluate correction methods via full-reference image quality assessment (FR-IQA) metrics. It provides properly exposed, overexposed, and underexposed images and associated subjective scores.

### 4. IEMX: A DCNN-Based IELA Model

In this section, we discuss how to build an IELA model. It is desired that such a model can accurately and efficiently predict the perceptual exposure level of a given image. Such a problem can be naturally formulated as a regression problem, which can be well addressed by DCNN (Deep Convolutional Neural Network) models [45] in an end-to-end manner, mapping from input images to their associated exposure levels.

As is widely known, in the last five or six years, thanks to the emergence and maturity of DCNN, the field of multimedia processing has developed rapidly. In essence, DCNN is a representation learning technology [45]. During training, by providing a large amount of raw data to the DCNN model, it can automatically discover the suitable internal representation of data. Today, in many technical fields, DCNN-based approaches usually perform much better than non-DCNN-based ones due to the availability of larger training sets, deeper models, better training algorithms, and more powerful GPUs. The first CNN was invented by LeCun in 1989 [53], and since the year of 2012, more elegant and powerful DCNN architectures have been continuously proposed in the literature, such as AlexNet [54], VGG [55], GoogLeNet [56], ResNet [57], DenseNet [58], and ShuffleNet [59].

We denote the proposed DCNN-based IELA model by IEMX, where “IELA” is short for “Image Exposure Metric” and “X” represents the concrete DCNN model used (in this paper, four specific DCNN models are investigated in the framework of IEMX, including GoogLeNet [56], ResNet [57], DenseNet [58], and ShuffleNet [59]). For training IEMX, the established dataset IEpsD_S described in Section 3 is used. The loss function is defined as

$$L(W) = \frac{1}{N} \sum_{j=1}^{N} \| IEMX(W; I_j) - s_j \|^2_F + \frac{\lambda}{2} \| W \|^2_F,$$ \hspace{1cm} (9)

where $W$ denotes the weights of the network, $\lambda$ is a regularization parameter, $I_j$ is the $j$th training image whose subjective exposure level is $s_j$, $\| W \|^2_F$ returns $W$’s Frobenius norm, and $N$ is the number of training samples in IEpsD_S. Implementation details of IEMX are presented in Section 5.1.

### 5. Experimental Results and Discussion

#### 5.1. Implementation Details of IEMX

Four state-of-the-art or representative DCNN architectures, including GoogLeNet [56], ResNet [57], DenseNet [58], and ShuffleNet [59], are investigated in the framework of IEMX, and the corresponding concrete IELA models are referred to as IEMG, IEMR, IEMD, and IEMS, respectively.

IEMXS were trained on IEpsD_S. For training IEMX, we used the fine-tuning strategy, i.e., IEMX was fine-tuned from the deep model pretrained on ImageNet [60] for the task of image classification. Actually, for GoogLeNet, ResNet, DenseNet, and ShuffleNet, the models pretrained on ImageNet were provided by their authors and we used them directly in IEMX. TensorFlow [61] was used as our deep learning platform. Key hyperparameters used when training IEMXS were set as “optimizer” = “ADAM” [62], “learning rate” = 0.001, “batch size” = 8, and “weight decay” = 0.0001.

#### 5.2. Test Protocol

The collected dataset IEpsD_R was used to evaluate the approaches’ capability for predicting the image’s perceptual exposure level. The performance of representative blind NR-IQA models, QA models for contrast-distorted images, and models specially designed for IELA was thoroughly studied and analyzed.

Four widely accepted metrics are adopted to evaluate the performance of the competing methods. The first two are the Spearman rank-order correlation coefficient (SROCC) and the Kendall rank-order correlation coefficient (KROCC). Both of them compute the correlation coefficients between the objective scores predicted by the IELA models and the
subjective exposure scores provided by the dataset. SROCC is defined as

\[
\text{SROCC} = 1 - \frac{6 \sum_{i=1}^{M} d_i^2}{M(M^2 - 1)}
\]  

(10)

where \(d_i\) is the difference between \(i\)th image’s ranks in objective and subjective judgements and \(M\) is the number of images in the test set. KROCC is defined as

\[
\text{KROCC} = \frac{M_c - M_d}{1/2M(M - 1)}
\]

(11)

where \(M_c\) and \(M_d\) are the numbers of concordant and discordant pairs in the test set, respectively. SROCC and KROCC are both nonparametric rank-based correlation metrics, implying that they depend only on the rank of the data points.

The third metric is the Pearson linear correlation coefficient (PLCC) between subjective scores and objective scores after a nonlinear mapping. Denote by \(\{s_i\}_{i=1}^{M}\) and \(\{o_i\}_{i=1}^{M}\) the set of subjective scores and the set of corresponding objective scores, respectively. First, a nonlinear mapping given by the following regression function [20] is applied to \(o_i\):

\[
q_i = f(o_i) = \beta_1 \left( 1 - \frac{1}{1 + \exp(\beta_2 (o_i - \beta_3))} \right) + \beta_4 o_i + \beta_5,
\]

(12)

where \(\beta_i, i = 1, 2, \ldots, 5\) are the model parameters that could be fitted using a nonlinear regression process to maximize the correlation between \(\{q_i\}_{i=1}^{M}\) and \(\{s_i\}_{i=1}^{M}\). After that, we can compute the PLCC value by

\[
\text{PLCC} = \frac{\sum_{i=1}^{M} (q_i - \bar{q})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^{M} (q_i - \bar{q})^2} \sqrt{\sum_{i=1}^{M} (o_i - \bar{o})^2}}
\]

(13)

The last metric is the RMSE (root mean squared error) between \(\{s_i\}_{i=1}^{M}\) and \(\{q_i\}_{i=1}^{M}\), which is defined as

\[
\text{RMSE} = \sqrt{\frac{1}{M} \sum_{i=1}^{M} (s_i - q_i)^2}.
\]

Different from SROCC and KROCC, PLCC and RMSE can measure the prediction accuracy of IELA models.

A better IELA model is anticipated to have higher SROCC, KROCC, and PLCC values and a lower RMSE value.

5.3. Evaluations of QA Models for Contrast-Distorted Images. As we have stated in Section 2.2, in most cases, improper exposure can decrease the image’s contrast. Thus, the studies focusing on quality assessment of contrast-distorted images (QACDI) are quite relevant to our work and it is reasonable to clearly know their performance for addressing the problem of IELA. Therefore, in this experiment, we evaluated the performance of six eminent QACDI models on IEpsD_R. The QACDI models evaluated included logAME [17], NR-CDIQA [12], NIQMC [18], and methods in [11, 13, 14]. It needs to be noted that NR-CDIQA and models in [11, 13, 14] are based on supervised learning and they were trained on a subset of CSIQ [21] comprising contrast-distorted images with associated subjective quality scores. The evaluation results are listed in Table 2. In addition, we also list the results of our IELA model IEMSN in Table 2 for comparison.

5.4. Evaluations of Blind NR-IQA Models. Blind NR-IQA is a research area closely related to IELA. In order to clearly know whether the existing blind NR-IQA models could address the IELA problem well, in this experiment, we evaluated the performance of several prominent blind NR-IQA models on IEpsD_R. These models include BRISQUE [31], SSEQ [28], OG-IQA [27], NOREQI [32], CS-BIQA [33], HyperIQA [63], NIQE [36], QAC [38], IL-NIQUE [37], LPSI [40], and TCLT [42].

Actually, existing blind NR-IQA models can be classified into two categories, opinion-aware ones and opinion-unaware ones. The opinion-aware models are obtained by training on a dataset comprising distorted images and associated subjective scores while the opinion-unaware ones do not require those kinds of training sets. BRISQUE [31], SSEQ [28], OG-IQA [27], NOREQI [32], CS-BIQA [33], and HyperIQA [63] are opinion-aware ones, and in this experiment, we used the trained models provided by their authors (for these five blind NR-IQA schemes, models provided by the authors were all trained on the entire LIVE dataset [20]). The other five models, NIQE [36], QAC [38], 

![Figure 8](image-url)
5.5. Evaluations of IELA Models. In this experiment, the performance of approaches specially designed for IELA was evaluated. The investigated approaches included Liu et al.’s approach [7], Rychagov and Efimov’s approach [8], Romaniak et al.’s approach [4], IEMCN, IEMRN, IEMDN, and IEMSN. The latter four are the four concrete forms of our proposed IELA model IEMX. The evaluation results are reported in Table 4.

In order to make a more convincing conclusion on the performance of the models, some statistical analysis is necessary [64]. We performed a left-tailed F-test [65] based on the prediction residuals of each model. The results of the significance test are shown in Figure 9. It can be seen that our method is much better than all other models.

5.6. Discussion. Based on the experimental results reported in Sections 5.3 ~ 5.5, the following conclusions could be drawn.

(1) Existing QACDI models or blind NR-IQA models cannot address the problem of IELA quite well. From the results presented in Tables 2 and 3, it can be seen that using these models, the assessment results of images’ exposure levels do not correlate well with the subjective evaluations. Specifically, the best-performing QACDI model is Xu and Wang’s model [14], whose SROCC value is 0.6716, and the best-performing blind NR-IQA model is QAC, whose SROCC value is 0.5415. Both of them perform much worse than the approaches specially designed for IELA, whose results are reported in Table 4.

The poor performance of blind NR-IQA algorithms and QACDI models should be mainly attributed to the fact that they cannot tell whether the quality distortion is caused by overexposure or underexposure. Another reason is that none of the existing datasets commonly used to train IQA models comprises image samples with associated subjective exposure level scores.

| Table 2: Performance evaluation of QA models for contrast-distorted images on IEpsD_R. |
| Methods | SROCC | KROCC | PLCC | RMSE |
| logAME [17] | 0.5312 | 0.3727 | 0.6191 | 0.4256 |
| NR-CDIQA [12] | 0.2794 | 0.2367 | 0.3139 | 0.5146 |
| NIQMC [18] | −0.0399 | −0.0032 | 0.4306 | 0.4892 |
| Ahmed et al. [11] | −0.1612 | −0.0529 | 0.4730 | 0.4775 |
| Wu et al. [13] | −0.0194 | −0.0036 | 0.1091 | 0.5388 |
| Xu and Wang [14] | 0.6716 | 0.4782 | 0.7214 | 0.3753 |
| IEMSN | 0.9850 | 0.8876 | 0.9750 | 0.1047 |

IL-NIQE [37], LPSI [40], and TCLT [42], are opinion-unaware ones.

Results of this experiment are reported in Table 3. In addition, we also list the results of our IELA model IEMSN in Table 3 for comparison.

| Table 3: Performance evaluation of blind NR-IQA models on IEpsD_R. |
| Methods | SROCC | KROCC | PLCC | RMSE |
| BRISQUE | 0.1799 | 0.0887 | 0.5148 | 0.4647 |
| NIQE | −0.0566 | −0.0395 | 0.2428 | 0.5258 |
| QAC | 0.5415 | 0.3976 | 0.5960 | 0.4352 |
| IL-NIQE | 0.4399 | 0.2939 | 0.5484 | 0.4532 |
| SSEQ | 0.3565 | 0.2321 | 0.5264 | 0.4608 |
| LPSI | 0.4945 | 0.3510 | 0.5514 | 0.4521 |
| TCLT | 0.1425 | 0.0963 | 0.3081 | 0.5156 |
| OG-IQA | −0.0124 | −0.0066 | 0.1005 | 0.5393 |
| NOREQI | 0.4022 | 0.2912 | 0.4117 | 0.4938 |
| CS-BIQA | 0.4352 | 0.3147 | 0.4905 | 0.4722 |
| HyperIQA | 0.4247 | 0.3175 | 0.5132 | 0.4567 |
| IEMSN | 0.9850 | 0.8876 | 0.9750 | 0.1047 |

The proposed DCNN-based IELA model IEMX performs extremely well for predicting perceptual exposure levels of real-world images. From the results listed in Table 4, it can be seen that all the variants of IEMX can achieve high SROCC, KROCC, and PLCC values and low RMSE values. IEMX’s performance is greatly better than the other IELA models evaluated for comparison. Especially, IEMSN performs the best among all the models evaluated, whose SROCC and PLCC values are 0.9850 and 0.9750, respectively.

(2) The proposed DCNN-based IELA model IEMX performs extremely well for predicting perceptual exposure levels of real-world images. From the results listed in Table 4, it can be seen that all the variants of IEMX can achieve high SROCC, KROCC, and PLCC values and low RMSE values. IEMX’s performance is greatly better than the other IELA models evaluated for comparison. Especially, IEMSN performs the best among all the models evaluated, whose SROCC and PLCC values are 0.9850 and 0.9750, respectively.

(3) The proposed method for generating synthetic images with various exposure levels is quite reasonable. In order to provide sufficient data for training the DCNN-based IELA model IEMX, we propose a method to generate synthetic images with various exposure levels as described in Section 3.2. With this strategy, we generated the dataset IEpsD_S, based on which IEMX was trained. Then, IEMX was tested on IEpsD_R, consisting of real-world images. In other words, IEMX was trained on synthetic images, but it was tested on real-world images. The results reported in Table 4 demonstrate that even though IEMX was trained on synthetic data, they perform quite well in predicting a real-world image’s exposure level. This fact implies that the scheme we proposed for generating synthetic images with various exposure levels is quite effective. Such a scheme significantly reduces the cost of preparing data for training IELA models. How to effectively make use of synthetic data to solve
vision problems should be given more attention by researchers.

6. Conclusion and Future Work

IELA models are highly desired in some vision-related industries. However, systematic studies specially focusing on this issue are still lacking. This work attempts to fill this research gap, and the contributions are from two aspects. First, an Image Exposure Database, namely, IEpsD, containing 24,500 images with multiple exposure levels, was established. For each image in IEpsD, we provide it a subjective exposure score representing its perceptual exposure level. IEpsD can serve as a benchmark to train and test IELA models. To the best of our knowledge, IEpsD is the first of its kind. Second, we formulated the IELA problem as a regression problem and proposed a DCNN-based solution IEMX. Four specific DCNN architectures, GoogLeNet, ResNet, DenseNet, and ShuffleNet, were investigated in the framework of IEMX. Experimental results show that IEMX yields much better exposure level prediction performance than all the compared competing methods. Experimental results also corroborate that blind NR-IQA models or QACDI models could not yield acceptable performance when being exploited to address the IELA issue. In near future, we will consider how to embed IELA metrics into the design of autoexposure algorithms.

Data Availability

The relevant source code and dataset have been made publicly available at https://cslinzhang.github.io/imgExpo/.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This study was supported in part by the National Natural Science Foundation of China under grant nos. 61973235, 61936014, and 61972285, in part by the Natural Science Foundation of Shanghai under grant no. 19ZR1461300, and in part by the Shanghai Science and Technology Innovation Plan under grant no. 20510760400.

References


