Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement

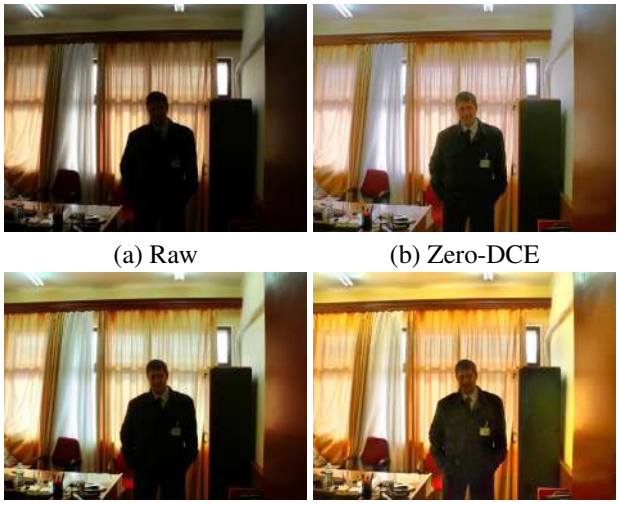
Chunle Guo[1,2*] Chongyi Li[1,2*] Jichang Guo[1†]
Chen Change Loy[3] Junhui Hou[2] Sam Kwong[2] Runmin Cong[4]
[1] BIIT Lab, Tianjin University [2] City University of Hong Kong [3] Nanyang Technological University [4] Beijing Jiaotong University
{guochunle,lichongyi,jcguo}@tju.edu.cn ccloy@ntu.edu.sg
{jh.hou,cssamk}@cityu.edu.hk rmcong@bjtu.edu.cn
https://li-chongyi.github.io/Proj_Zero-DCE.html/

Abstract

The paper presents a novel method, Zero-Reference Deep Curve Estimation (Zero-DCE), which formulates light enhancement as a task of image-specific curve estimation with a deep network. Our method trains a lightweight deep network, DCE-Net, to estimate pixel-wise and high-order curves for dynamic range adjustment of a given image. The curve estimation is specially designed, considering pixel value range, monotonicity, and differentiability. Zero-DCE is appealing in its relaxed assumption on reference images, i.e., it does not require any paired or unpaired data during training. This is achieved through a set of carefully formulated non-reference loss functions, which implicitly measure the enhancement quality and drive the learning of the network. Our method is efficient as image enhancement can be achieved by an intuitive and simple nonlinear curve mapping. Despite its simplicity, we show that it generalizes well to diverse lighting conditions. Extensive experiments on various benchmarks demonstrate the advantages of our method over state-of-the-art methods qualitatively and quantitatively. Furthermore, the potential benefits of our Zero-DCE to face detection in the dark are discussed.

1. Introduction

Many photos are often captured under suboptimal lighting conditions due to inevitable environmental and/or technical constraints. These include inadequate and unbalanced lighting conditions in the environment, incorrect placement of objects against extreme back light, and under-exposure during image capturing. Such low-light photos suffer from compromised aesthetic quality and unsatisfactory transmission of information. The former affects viewers' experience while the latter leads to wrong message being communicated, such as inaccurate object/face recognition.



(a) Raw (b) Zero-DCE
(c) Wang *et al*. [28] (d) EnlightenGAN [12]

Figure 1: Visual comparisons on a typical low-light image. The proposed Zero-DCE achieves visually pleasing result in terms of brightness, color, contrast, and naturalness, while existing methods either fail to cope with the extreme back light or generate color artifacts. In contrast to other deep learning-based methods, our approach is trained without any reference image.

In this study, we present a novel deep learning-based method, Zero-Reference Deep Curve Estimation (Zero-DCE), for low-light image enhancement. It can cope with diverse lighting conditions including nonuniform and poor lighting cases. Instead of performing image-to-image mapping, we reformulate the task as an image-specific curve estimation problem. In particular, the proposed method takes a low-light image as input and produces high-order curves as its output. These curves are then used for pixel-wise adjustment on the dynamic range of the input to obtain an enhanced image. The curve estimation is carefully formulated so that it maintains the range of the enhanced image and preserves the contrast of neighboring pixels. Importantly, it

*The first two authors contribute equally to this work.
†Jichang Guo (jcguo@tju.edu.cn) is the corresponding author.

is differentiable, and thus we can learn the adjustable parameters of the curves through a deep convolutional neural network. The proposed network is lightweight and it can be iteratively applied to approximate higher-order curves for more robust and accurate dynamic range adjustment.

A unique advantage of our deep learning-based method is **zero-reference**, *i.e.*, it does not require any paired or even unpaired data in the training process as in existing CNN-based [28,32] and GAN-based methods [12,38]. This is made possible through a set of specially designed non-reference loss functions including spatial consistency loss, exposure control loss, color constancy loss, and illumination smoothness loss, all of which take into consideration multi-factor of light enhancement. We show that even with zero-reference training, Zero-DCE can still perform competitively against other methods that require paired or unpaired data for training. An example of enhancing a low-light image comprising nonuniform illumination is shown in Fig. 1. Comparing to state-of-the-art methods, Zero-DCE brightens up the image while preserving the inherent color and details. In contrast, both CNN-based method [28] and GAN-based EnlightenGAN [12] yield under-(the face) and over-(the cabinet) enhancement.

Our **contributions** are summarized as follows.

1) We propose the first low-light enhancement network that is independent of paired and unpaired training data, thus avoiding the risk of overfitting. As a result, our method generalizes well to various lighting conditions.

2) We design an image-specific curve that is able to approximate pixel-wise and higher-order curves by iteratively applying itself. Such image-specific curve can effectively perform mapping within a wide dynamic range.

3) We show the potential of training a deep image enhancement model in the absence of reference images through task-specific non-reference loss functions that indirectly evaluate enhancement quality.

Our Zero-DCE method supersedes state-of-the-art performance both in qualitative and quantitative metrics. More importantly, it is capable of improving high-level visual tasks, *e.g.*, face detection, without inflicting high computational burden. **It is capable of processing images in real-time (about 500 FPS for images of size 640×480×3 on GPU) and takes only 30 minutes for training.**

2. Related Work

Conventional Methods. HE-based methods perform light enhancement through expanding the dynamic range of an image. Histogram distribution of images is adjusted at both global [7, 10] and local levels [15, 27]. There are also various methods adopting the Retinex theory [13] that typically decomposes an image into reflectance and illumination. The reflectance component is commonly assumed to be consistent under any lighting conditions; thus, light enhancement is formulated as an illumination estimation problem. Building on the Retinex theory, several methods have been proposed. Wang *et al*. [29] designed a naturalness- and information-preserving method when handling images of nonuniform illumination; Fu *et al*. [8] proposed a weighted variation model to simultaneously estimate the reflectance and illumination of an input image; Guo *et al*. [9] first estimated a coarse illumination map by searching the maximum intensity of each pixel in RGB channels, then refining the coarse illumination map by a structure prior; Li *et al*. [19] proposed a new Retinex model that takes noise into consideration. The illumination map was estimated through solving an optimization problem. Contrary to the conventional methods that fortuitously change the distribution of image histogram or that rely on potentially inaccurate physical models, the proposed Zero-DCE method produces an enhanced result through image-specific curve mapping. Such a strategy enables light enhancement on images without creating unrealistic artifacts. Yuan and Sun [36] proposed an automatic exposure correction method, where the S-shaped curve for a given image is estimated by a global optimization algorithm and each segmented region is pushed to its optimal zone by curve mapping. Different from [36], our Zero-DCE is a purely data-driven method and takes multiple light enhancement factors into consideration in the design of the non-reference loss functions, and thus enjoys better robustness, wider image dynamic range adjustment, and lower computational burden.

Data-Driven Methods. Data-driven methods are largely categorized into two branches, namely CNN-based and GAN-based methods. Most CNN-based solutions rely on paired data for supervised training, therefore they are resource-intensive. Often time, the paired data are exhaustively collected through automatic light degradation, changing the settings of cameras during data capturing, or synthesizing data via image retouching. For example, the LLNet [20] was trained on data simulated on random Gamma correction; the LOL dataset [32] of paired low/normal light images was collected through altering the exposure time and ISO during image acquisition; the MIT-Adobe FiveK dataset [3] comprises 5,000 raw images, each of which has five retouched images produced by trained experts.

Recently, Wang *et al*. [28] proposed an underexposed photo enhancement network by estimating the illumination map. This network was trained on paired data that were retouched by three experts. Understandably, light enhancement solutions based on paired data are impractical in many ways, considering the high cost involved in collecting sufficient paired data as well as the inclusion of factitious and unrealistic data in training the deep models. Such constraints are reflected in the poor generalization capability of CNN-based methods. Artifacts and color casts are com-

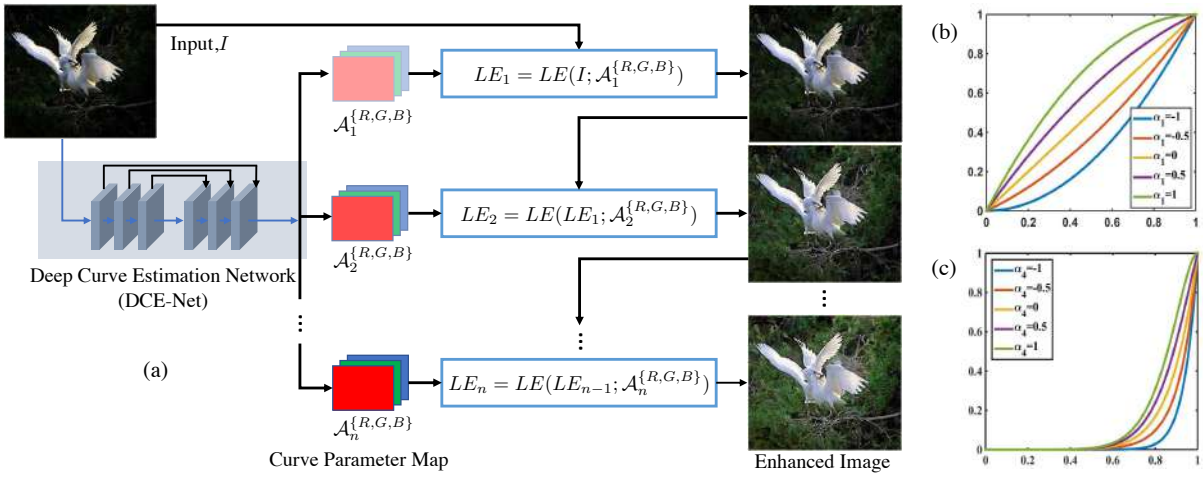


Figure 2: (a) The framework of Zero-DCE. A DCE-Net is devised to estimate a set of best-fitting Light-Enhancement curves (LE-curves) that iteratively enhance a given input image. (b, c) LE-curves with different adjustment parameters α and numbers of iteration n. In (c), α_1, α_2, and α_3 are equal to -1 while n is equal to 4. In each subfigure, the horizontal axis represents the input pixel values while the vertical axis represents the output pixel values.

monly generated, when these methods are presented with real-world images of various light intensities.

Unsupervised GAN-based methods have the advantage of eliminating paired data for training. EnlightenGAN [12], an unsupervised GAN-based and pioneer method that learns to enhance low-light images using unpaired low/normal light data. The network was trained by taking into account elaborately designed discriminators and loss functions. However, unsupervised GAN-based solutions usually require careful selection of unpaired training data.

The proposed Zero-DCE is superior to existing data-driven methods in three aspects. First, it explores a new learning strategy, *i.e.*, one that requires *zero reference*, hence eliminating the need for paired and unpaired data. Second, the network is trained by taking into account carefully defined non-reference loss functions. This strategy allows output image quality to be implicitly evaluated, the results of which would be reiterated for network learning. Third, our method is highly efficient and cost-effective. These advantages benefit from our zero-reference learning framework, lightweight network structure, and effective non-reference loss functions.

3. Methodology

We present the framework of Zero-DCE in Fig. 2. A Deep Curve Estimation Network (DCE-Net) is devised to estimate a set of best-fitting Light-Enhancement curves (LE-curves) given an input image. The framework then maps all pixels of the input's RGB channels by applying the curves iteratively for obtaining the final enhanced image. We next detail the key components in Zero-DCE, namely LE-curve, DCE-Net, and non-reference loss functions in the following sections.

3.1. Light-Enhancement Curve (LE-curve)

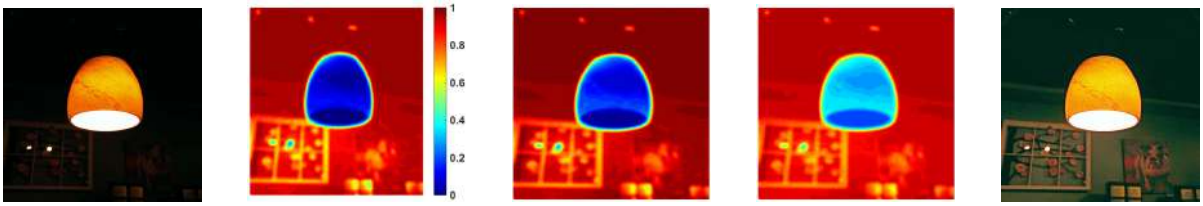
Inspired by the curves adjustment used in photo editing software, we attempt to design a kind of curve that can map a low-light image to its enhanced version automatically, where the self-adaptive curve parameters are solely dependent on the input image. There are three objectives in the design of such a curve: 1) each pixel value of the enhanced image should be in the normalized range of [0,1] to avoid information loss induced by overflow truncation; 2) this curve should be monotonous to preserve the differences (contrast) of neighboring pixels; and 3) the form of this curve should be as simple as possible and differentiable in the process of gradient backpropagation.

To achieve these three objectives, we design a quadratic curve, which can be expressed as:

$$LE(I(\mathbf{x}); \alpha) = I(\mathbf{x}) + \alpha I(\mathbf{x})(1 - I(\mathbf{x})), \qquad (1)$$

where $\mathbf{x}$ denotes pixel coordinates, $LE(I(\mathbf{x}); \alpha)$ is the enhanced version of the given input $I(\mathbf{x})$, $\alpha \in [-1, 1]$ is the trainable curve parameter, which adjusts the magnitude of LE-curve and also controls the exposure level. Each pixel is normalized to $[0, 1]$ and all operations are pixel-wise. We separately apply the LE-curve to three RGB channels instead of solely on the illumination channel. The three-channel adjustment can better preserve the inherent color and reduce the risk of over-saturation. We report more details in the supplementary material.

An illustration of LE-curves with different adjustment parameters α is shown in Fig. 2(b). It is clear that the LE-curve complies with the three aforementioned objectives. In addition, the LE-curve enables us to increase or decrease the dynamic range of an input image. This capability is conducive to not only enhancing low-light regions but also



(a) Input (b) $\mathcal{A}_n^R$ (c) $\mathcal{A}_n^G$ (d) $\mathcal{A}_n^B$ (e) Result

Figure 3: An example of the pixel-wise curve parameter maps. For visualization, we average the curve parameter maps of all iterations ($n = 8$) and normalize the values to the range of $[0, 1]$. $\mathcal{A}_n^R$, $\mathcal{A}_n^G$, and $\mathcal{A}_n^B$ represent the averaged best-fitting curve parameter maps of R, G, and B channels, respectively. The maps in (b), (c), and (d) are represented by heatmaps.

removing over-exposure artifacts.

Higher-Order Curve. The LE-curve defined in Eq. (1) can be applied iteratively to enable more versatile adjustment to cope with challenging low-light conditions. Specifically,

$$LE_n(\mathbf{x}) = LE_{n-1}(\mathbf{x}) + \alpha_n LE_{n-1}(\mathbf{x})(1 - LE_{n-1}(\mathbf{x})), \tag{2}$$

where n is the number of iteration, which controls the curvature. In this paper, we set the value of n to 8, which can deal with most cases satisfactory. Eq. (2) can be degraded to Eq. (1) when n is equal to 1. Figure 2(c) provides an example showing high-order curves with different α and n, which have more powerful adjustment capability (*i.e.*, greater curvature) than the curves in Figure 2(b).

Pixel-Wise Curve. A higher-order curve can adjust an image within a wider dynamic range. Nonetheless, it is still a global adjustment since α is used for all pixels. A global mapping tends to over-/under- enhance local regions. To address this problem, we formulate α as a pixel-wise parameter, *i.e.*, each pixel of the given input image has a corresponding curve with the best-fitting α to adjust its dynamic range. Hence, Eq. (2) can be reformulated as:

$$LE_n(\mathbf{x}) = LE_{n-1}(\mathbf{x}) + \mathcal{A}_n(\mathbf{x}) LE_{n-1}(\mathbf{x})(1 - LE_{n-1}(\mathbf{x})), \tag{3}$$

where $\mathcal{A}$ is a parameter map with the same size as the given image. Here, we assume that pixels in a local region have the same intensity (also the same adjustment curves), and thus the neighboring pixels in the output result still preserve the monotonous relations. In this way, the pixel-wise higher-order curves also comply with three objectives.

We present an example of the estimated curve parameter maps of three channels in Fig. 3. As shown, the best-fitting parameter maps of different channels have similar adjustment tendency but different values, indicating the relevance and difference among the three channels of a low-light image. The curve parameter map accurately indicates the brightness of different regions (*e.g.*, the two glitters on the wall). With the fitting maps, the enhanced version image can be directly obtained by pixel-wise curve mapping. As shown in Fig. 3(e), the enhanced version reveals the content in dark regions and preserves the bright regions.

3.2. DCE-Net

To learn the mapping between an input image and its best-fitting curve parameter maps, we propose a Deep Curve Estimation Network (DCE-Net). The input to the DCE-Net is a low-light image while the outputs are a set of pixel-wise curve parameter maps for corresponding higher-order curves. We employ a plain CNN of seven convolutional layers with symmetrical concatenation. Each layer consists of 32 convolutional kernels of size 3×3 and stride 1 followed by the ReLU activation function. We discard the down-sampling and batch normalization layers that break the relations of neighboring pixels. The last convolutional layer is followed by the Tanh activation function, which produces 24 parameter maps for 8 iterations ($n = 8$), where each iteration requires three curve parameter maps for the three channels. The detailed architecture of DCE-Net is provided in the supplementary material. It is noteworthy that DCE-Net only has 79,416 trainable parameters and 5.21G Flops for an input image of size 256×256×3. It is therefore lightweight and can be used in computational resource-limited devices, such as mobile platforms.

3.3. Non-Reference Loss Functions

To enable zero-reference learning in DCE-Net, we propose a set of differentiable non-reference losses that allow us to evaluate the quality of enhanced images. The following four types of losses are adopted to train our DCE-Net.

Spatial Consistency Loss. The spatial consistency loss L_{spa} encourages spatial coherence of the enhanced image through preserving the difference of neighboring regions between the input image and its enhanced version:

$$L_{spa} = \frac{1}{K} \sum_{i=1}^{K} \sum_{j \in \Omega(i)} (|(Y_i - Y_j)| - |(I_i - I_j)|)^2, \tag{4}$$

where K is the number of local region, and $\Omega(i)$ is the four neighboring regions (top, down, left, right) centered at the region i. We denote Y and I as the average intensity value of the local region in the enhanced version and input image, respectively. We empirically set the size of the local region to 4×4. This loss is stable given other region sizes.

(a) Input (b) Zero-DCE (c) w/o L_{spa} (d) w/o L_{exp} (e) w/o L_{col} (f) w/o $L_{tv_{\mathcal{A}}}$

Figure 4: Ablation study of the contribution of each loss (spatial consistency loss L_{spa}, exposure control loss L_{exp}, color constancy loss L_{col}, illumination smoothness loss $L_{tv_{\mathcal{A}}}$).

Exposure Control Loss. To restrain under-/over-exposed regions, we design an exposure control loss L_{exp} to control the exposure level. The exposure control loss measures the distance between the average intensity value of a local region to the well-exposedness level E. We follow existing practices [23, 24] to set E as the gray level in the RGB color space. We set E to 0.6 in our experiments although we do not find much performance difference by setting E within $[0.4, 0.7]$. The loss L_{exp} can be expressed as:

$$L_{exp} = \frac{1}{M}\sum_{k=1}^{M}|Y_k - E|, \tag{5}$$

where M represents the number of nonoverlapping local regions of size 16×16, Y is the average intensity value of a local region in the enhanced image.

Color Constancy Loss. Following Gray-World color constancy hypothesis [2] that color in each sensor channel averages to gray over the entire image, we design a color constancy loss to correct the potential color deviations in the enhanced image and also build the relations among the three adjusted channels. The color constancy loss L_{col} can be expressed as:

$$L_{col} = \sum_{\forall(p,q)\in\varepsilon}(J^p - J^q)^2, \varepsilon = \{(R,G),(R,B),(G,B)\}, \tag{6}$$

where J^p denotes the average intensity value of p channel in the enhanced image, (p,q) represents a pair of channels.

Illumination Smoothness Loss. To preserve the monotonicity relations between neighboring pixels, we add an illumination smoothness loss to each curve parameter map $\mathcal{A}$. The illumination smoothness loss $L_{tv_{\mathcal{A}}}$ is defined as:

$$L_{tv_{\mathcal{A}}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{c\in\xi}(|\nabla_x\mathcal{A}_n^c| + \nabla_y\mathcal{A}_n^c|)^2, \xi = \{R,G,B\}, \tag{7}$$

where N is the number of iteration, ∇_x and ∇_y represent the horizontal and vertical gradient operations, respectively.

Total Loss. The total loss can be expressed as:

$$L_{total} = L_{spa} + L_{exp} + W_{col}L_{col} + W_{tv_{\mathcal{A}}}L_{tv_{\mathcal{A}}}, \tag{8}$$

where W_{col} and $W_{tv_{\mathcal{A}}}$ are the weights of the losses.

4. Experiments

Implementation Details. CNN-based models usually use self-captured paired data for network training [5, 17, 28, 30, 32, 33] while GAN-based models elaborately select unpaired data [6, 11, 12, 16, 35]. To bring the capability of wide dynamic range adjustment into full play, we incorporate both low-light and over-exposed images into our training set. To this end, we employ 360 multi-exposure sequences from the Part1 of SICE dataset [4] to train the proposed DCE-Net. The dataset is also used as a part of the training data in EnlightenGAN [12]. We randomly split 3,022 images of different exposure levels in the Part1 subset [4] into two parts (2,422 images for training and the rest for validation). We resize the training images to the size of 512×512.

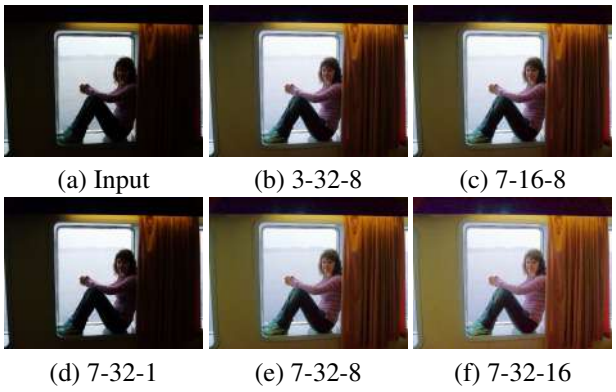
We implement our framework with PyTorch on an NVIDIA 2080Ti GPU. A batch size of 8 is applied. The filter weights of each layer are initialized with standard zero mean and 0.02 standard deviation Gaussian function. Bias is initialized as a constant. We use ADAM optimizer with default parameters and fixed learning rate $1e^{-4}$ for our network optimization. The weights W_{col} and $W_{tv_{\mathcal{A}}}$ are set to 0.5, and 20, respectively, to balance the scale of losses.

4.1. Ablation Study

We perform several ablation studies to demonstrate the effectiveness of each component of Zero-DCE as follows. More qualitative and quantitative comparisons can be found in the supplementary material.

Contribution of Each Loss. We present the results of Zero-DCE trained by various combinations of losses in Fig. 4. The result without spatial consistency loss L_{spa} has relatively lower contrast (*e.g.*, the cloud regions) than the full result. This shows the importance of L_{spa} in preserving the difference of neighboring regions between the input and the enhanced image. Removing the exposure control loss L_{exp} fails to recover the low-light region. Severe color casts emerge when the color constancy loss L_{col} is discarded. This variant ignores the relations among three channels when curve mapping is applied. Finally, removing the illumination smoothness loss $L_{tv_{\mathcal{A}}}$ hampers the correlations between neighboring regions leading to obvious artifacts.

Effect of Parameter Settings. We evaluate the effect of



(a) Input (b) 3-32-8 (c) 7-16-8

(d) 7-32-1 (e) 7-32-8 (f) 7-32-16

Figure 5: Ablation study of the effect of parameter settings. l-f-n represents the proposed Zero-DCE with l convolutional layers, f feature maps of each layer (except the last layer), and n iterations.

parameters in Zero-DCE, consisting of the depth and width of the DCE-Net and the number of iterations. A visual example is presented in Fig. 5. In Fig. 5(b), with just three convolutional layers, Zero-DCE$_{3-32-8}$ can already produce satisfactory results, suggesting the effectiveness of zero-reference learning. The Zero-DCE$_{7-32-8}$ and Zero-DCE$_{7-32-16}$ produce most visually pleasing results with natural exposure and proper contrast. By reducing the number of iterations to 1, an obvious decrease in performance is observed on Zero-DCE$_{7-32-1}$ as shown in Fig. 5(d). This is because the curve with only single iteration has limited adjustment capability. This suggests the need for higher-order curves in our method. We choose Zero-DCE$_{7-32-8}$ as the final model based given its good trade-off between efficiency and restoration performance.

Impact of Training Data. To test the impact of training data, we retrain the Zero-DCE on different datasets: 1) only 900 low-light images out of 2,422 images in the original training set (Zero-DCE$_{Low}$), 2) 9,000 unlabeled low-light images provided in the DARK FACE dataset [37] (Zero-DCE$_{LargeL}$), and 3) 4800 multi-exposure images from the data augmented combination of Part1 and Part2 subsets in the SICE dataset [4] (Zero-DCE$_{LargeLH}$). As shown in Fig. 6(c) and (d), after removing the over-exposed training data, Zero-DCE tends to over-enhance the well-lit regions (*e.g.*, the face), in spite of using more low-light images, (*i.e.*, Zero-DCE$_{LargeL}$). Such results indicate the rationality and necessity of the usage of multi-exposure training data in the training process of our network. In addition, the Zero-DCE can better recover the dark regions when more multi-exposure training data are used (*i.e.*, Zero-DCE$_{LargeLH}$), as shown in Fig. 6(e). For a fair comparison with other deep learning-based methods, we use a comparable amount of training data with them although more training data can bring better visual performance to our approach.

4.2. Benchmark Evaluations

We compare Zero-DCE with several state-of-the-art methods: three conventional methods (SRIE [8], LIME [9], Li *et al*. [19]), two CNN-based methods (RetinexNet [32], Wang *et al*. [28]), and one GAN-based method (EnlightenGAN [12]). The results are reproduced by using publicly available source codes with recommended parameters.

We perform qualitative and quantitative experiments on standard image sets from previous works including NPE [29] (84 images), LIME [9] (10 images), MEF [22] (17 images), DICM [14] (64 images), and VV[‡] (24 images). Besides, we quantitatively validate our method on the Part2 subset of SICE dataset [4], which consists of 229 multi-exposure sequences and the corresponding reference image for each multi-exposure sequence. For a fair comparison, we only use the low-light images of Part2 subset [4] for testing, since baseline methods cannot handle over-exposed images well. Specifically, we choose the first three (resp. four) low-light images if there are seven (resp. nine) images in a multi-exposure sequence and resize all images to a size of 1200×900×3. Finally, we obtain 767 paired low/normal light images. We discard the low/normal light image dataset mentioned in [37], because the training datasets of RetinexNet [32] and EnlightenGAN [12] consist of some images from this dataset. Note that the latest paired training and testing dataset constructed in [28] are not publicly available. We did not use the MIT-Adobe FiveK dataset [3] as it is not primarily designed for under-exposed photos enhancement.

4.2.1 Visual and Perceptual Comparisons

We present the visual comparisons on typical low-light images in Fig. 7. For challenging back-lit regions (*e.g.*, the face in Fig. 7(a)), Zero-DCE yields natural exposure and clear details while SRIE [8], LIME [9], Wang *et al*. [28], and EnlightenGAN [12] cannot recover the face clearly. RetinexNet [32] produces over-exposed artifacts. In the second example featuring an indoor scene, our method enhances dark regions and preserves color of the input image simultaneously. The result is visually pleasing without obvious noise and color casts. In contrast, Li *et al*. [19] over-smoothes the details while other baselines amplify noise and even produce color deviation (*e.g.*, the color of wall).

We perform a user study to quantify the subjective visual quality of various methods. We process low-light images from the image sets (NPE, LIME, MEF, DICM, VV) by different methods. For each enhanced result, we display it on a screen and provide the input image as a reference. A total of 15 human subjects are invited to independently score the visual quality of the enhanced image. These subject-

[‡] https://sites.google.com/site/vonikakis/datasets

(a) Input (b) Zero-DCE (c) Zero-DCE$_{Low}$ (d) Zero-DCE$_{LargeL}$ (e) Zero-DCE$_{LargeLH}$

Figure 6: Ablation study on the impact of training data.

(a) Inputs (b) SRIE [8] (c) LIME [9] (d) Li *et al*. [19]

(e) RetinexNet [32] (f) Wang *et al*. [28] (g) EnlightenGAN [12] (h) Zero-DCE

Figure 7: Visual comparisons on typical low-light images. Red boxes indicate the obvious differences.

s are trained by observing the results from 1) whether the results contain over-/under-exposed artifacts or over-/under-enhanced regions; 2) whether the results introduce color deviation; and 3) whether the results have unnatural texture and obvious noise. The scores of visual quality range from 1 to 5 (worst to best quality). The average subjective scores for each image set are reported in Table 1. As summarized in Table 1, Zero-DCE achieves the highest average User Study (US) score for a total of 202 testing images from the above-mentioned image sets. For the MEF, DICM, and VV sets, our results are most favored by the subjects. In addition to the US score, we employ a non-reference perceptual index (PI) [1, 21, 25] to evaluate the perceptual quality. The PI metric is originally used to measure perceptual quality in image super-resolution. It has also been used to assess the performance of other image restoration tasks, such as image dehazing [26]. A lower PI value indicates better perceptual quality. The PI values are reported in Table 1 too. Similar to the user study, the proposed Zero-DCE is superior to other competing methods in terms of the average PI values.

4.2.2 Quantitative Comparisons

For full-reference image quality assessment, we employ the Peak Signal-to-Noise Ratio (PSNR,dB), Structural Similarity (SSIM) [31], and Mean Absolute Error (MAE) metrics to quantitatively compare the performance of different methods on the Part2 subset [4]. In Table 2, the proposed Zero-DCE achieves the best values under all cases, despite that it does not use any paired or unpaired training data. Zero-DCE is also computationally efficient, benefited from the simple curve mapping form and lightweight network structure. Table 3 shows the runtime[§] of different methods averaged on 32 images of size 1200×900×3. For conventional methods, only the codes of CPU version are available.

[§]Runtime is measured on a PC with an Nvidia GTX 2080Ti GPU and Intel I7 6700 CPU, except for Wang *et al*. [28], which has to run on GTX 1080Ti GPU.

Table 1: User study (US)↑/Perceptual index (PI)↓ scores on the image sets (NPE, LIME, MEF, DICM, VV). Higher US score indicates better human subjective visual quality while lower PI value indicates better perceptual quality. The best result is in red whereas the second best one is in blue under each case.

Method	NPE	LIME	MEF	DICM	VV	Average
SRIE [8]	3.65/2.79	3.50/2.76	3.22/2.61	3.42/3.17	2.80/3.37	3.32/2.94
LIME [9]	3.78/3.05	3.95/3.00	3.71/2.78	3.31/3.35	3.21/3.03	3.59/3.04
Li *et al.* [19]	3.80/3.09	3.78/3.02	2.93/3.61	3.47/3.43	2.87/3.37	3.37/3.72
RetinexNet [32]	3.30/3.18	2.32/3.08	2.80/2.86	2.88/3.24	1.96/2.95	2.58/3.06
Wang *et al.* [28]	3.83/2.83	3.82/2.90	3.13/2.72	3.44/3.20	2.95/3.42	3.43/3.01
EnlightenGAN [12]	3.90/2.96	3.84/2.83	3.75/2.45	3.50/3.13	3.17/4.71	3.63/3.22
Zero-DCE	3.81/2.84	3.80/2.76	4.13/2.43	3.52/3.04	3.24/3.33	3.70/2.88

Table 2: Quantitative comparisons in terms of full-reference image quality assessment metrics. The best result is in red whereas the second best one is in blue under each case.

Method	PSNR↑	SSIM↑	MAE↓
SRIE [8]	14.41	0.54	127.08
LIME [9]	16.17	0.57	108.12
Li *et al.* [19]	15.19	0.54	114.21
RetinexNet [32]	15.99	0.53	104.81
Wang *et al.* [28]	13.52	0.49	142.01
EnlightenGAN [12]	16.21	0.59	102.78
Zero-DCE	16.57	0.59	98.78

Table 3: Runtime (RT) comparisons (in second). The best result is in red whereas the second best one is in blue.

Method	RT	Platform
SRIE [8]	12.1865	MATLAB (CPU)
LIME [9]	0.4914	MATLAB (CPU)
Li *et al.* [19]	90.7859	MATLAB (CPU)
RetinexNet [32]	0.1200	TensorFlow (GPU)
Wang *et al.* [28]	0.0210	TensorFlow (GPU)
EnlightenGAN [12]	0.0078	PyTorch (GPU)
Zero-DCE	0.0025	PyTorch (GPU)

4.2.3 Face Detection in the Dark

We investigate the performance of low-light image enhancement methods on the face detection task under low-light conditions. Specifically, we use the latest DARK FACE dataset [37] that composes of 10,000 images taken in the dark. Since the bounding boxes of test set are not publicly available, we perform evaluation on the training and validation sets, which consists of 6,000 images. A state-of-the-art deep face detector, Dual Shot Face Detector (DSFD) [18], trained on WIDER FACE dataset [34], is used as the baseline model. We feed the results of different low-light image enhancement methods to the DSFD [18] and depict the precision-recall (P-R) curves in Fig. 8. Besides, we also compare the average precision (AP) by using the evaluation tool[¶] provided in DARK FACE dataset [37].

[¶]https://github.com/Ir1d/DARKFACE_eval_tools

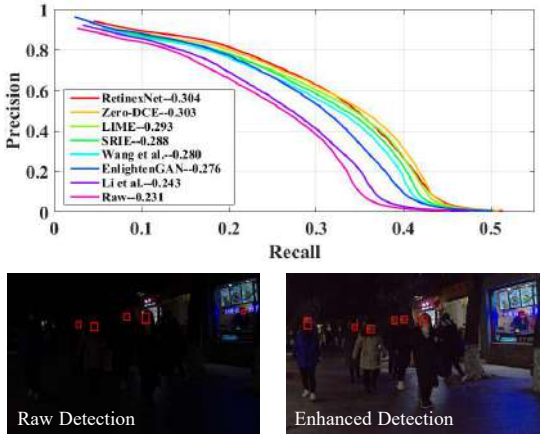


Figure 8: The performance of face detection in the dark. P-R curves, the AP, and two examples of face detection before and after enhanced by our Zero-DCE.

As shown in Fig. 8, after image enhancement, the precision of DSFD [18] increases considerably compared to that using raw images without enhancement. Among different methods, RetinexNet [32] and Zero-DCE perform the best. Both methods are comparable but Zero-DCE performs better in the high recall area. Observing the examples, our Zero-DCE lightens up the faces in the extremely dark regions and preserves the well-lit regions, thus improves the performance of face detector in the dark.

5. Conclusion

We proposed a deep network for low-light image enhancement. It can be trained end-to-end with zero reference images. This is achieved by formulating the low-light image enhancement task as an image-specific curve estimation problem, and devising a set of differentiable non-reference losses. Experiments demonstrate the superiority of our method against existing light enhancement methods. In future work, we will try to introduce semantic information to solve hard cases and consider the effects of noise.

Acknowledgements. This research was supported by NSFC (61771334,61632018,61871342), SenseTime-NTU Collaboration Project, Singapore MOE AcRF Tier 1 (2018-T1-002-056), NTU SUG, NTU NAP, Fundamental Research Funds for the Central Universities (2019RC039), China Postdoctoral Science Foundation (2019M660438), Hong Kong RGG (9048123) (CityU 21211518), Hong Kong GRF-RGC General Research Fund (9042322,9042489,9042816).

References

[1] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *CVPR*, 2018. 7

[2] Gershon Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Institute*, 310(1):1–26, 1980. 5

[3] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR*, 2011. 2, 6

[4] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure image. *IEEE Transactions on Image Processing*, 27(4):2049–2026, 2018. 5, 6, 7

[5] Chen Chen, Qifeng Chen, Jia Xu, and Koltun Vladlen. Learning to see in the dark. In *CVPR*, 2018. 5

[6] Yusheng Chen, Yuching Wang, Manhsin Kao, and Yungyu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *CVPR*, 2018. 5

[7] Dinu Coltuc, Philippe Bolon, and Jean-Marc Chassery. Exact histogram specification. *IEEE Transactions on Image Processing*, 15(5):1143–1152, 2006. 2

[8] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *CVPR*, 2016. 2, 6, 7, 8

[9] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017. 2, 6, 7, 8

[10] Haidi Ibrahim and Nicholas Sia Pik Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(4):1752–1758, 2007. 2

[11] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Wespe: Weakly supervised photo enhancer for digital cameras. In *CVPRW*, 2018. 5

[12] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. EnlightenGAN: Deep light enhancement without paired supervision. In *CVPR*, 2019. 1, 2, 3, 5, 6, 7, 8

[13] Edwin H Land. The retinex theory of color vision. *Scientific American*, 237(6):108–128, 1977. 2

[14] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation. In *ICIP*, 2012. 6

[15] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE Transactions on Image Processing*, 22(12):5372–5384, 2013. 2

[16] Chongyi Li, Chunle Guo, and Jichang Guo. Underwater image color correction based on weakly supervised color transfer. *IEEE Signal Processing Letters*, 25(3):323–327, 2018. 5

[17] Chongyi Li, Jichang Guo, Fatih Porikli, and Yanwei Pang. Lightennet: a convolutional neural network for weakly illuminated image enhancement. *Pattern Recognition Letters*, 104:15–22, 2018. 5

[18] Jian Li, Yabiao Wang, Changan Wang, Ying Tai, Jianjun Qian, Jian Yang, Chengjie Wang, Jilin Li, and Feiyuen Huang. Dsfd: Dual shot face detector. In *CVPR*, 2019. 8

[19] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 2, 6, 7, 8

[20] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 2

[21] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158:1–16, 2017. 7

[22] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015. 6

[23] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *PCCGA*, 2007. 5

[24] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion: A simple and practrical alterrnative to high dynamic range photography. *Computer Graphics Forum*, 28(1):161–171, 2009. 5

[25] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013. 7

[26] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *CVPR*, 2019. 7

[27] J Alex Stark. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Transactions on Image Processing*, 9(5):889–896, 2000. 2

[28] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *CVPR*, 2019. 1, 2, 5, 6, 7, 8

[29] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013. 2, 6

[30] Wenguan Wang, Qiuxia Lai, Huazhu Fu, Jianbing Shen, and Haibin Ling. Salient object detection in the deep learning era: An in-depth survey, 2019. 5

[31] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 7

[32] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *BMVC*, 2018. 2, 5, 6, 7, 8

[33] Peng Xu. Deep learning for free-hand sketch: A survey, 2020. 5

[34] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In *CVPR*, 2016. 8

[35] Runsheng Yu, Wenyu Liu, Yasen Zhang, Zhi Qu, Deli Zhao, and Bo Zhang. Deepexposure: Learning to expose photo-

s with asynchronously reinforced adversarial learning. In *NeurIPS*, 2018. 5

[36] Lu Yuan and Jian Sun. Automatic exposure correction of consumer photographs. In *ECCV*, 2012. 2

[37] Ye Yuan, Wenhan Yang, Wenqi Ren, Jiaying Liu, Walter J Scheirer, and Wang Zhangyang. Ug+ track 2: A collective benchmark effort for evaluating and advancing image understanding in poor visibility environments, 2019. arXiv arXiv:1904.04474. 6, 8

[38] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017. 2