Zero-Shot Restoration of Back-lit Images Using Deep Internal Learning

Lin Zhang
School of Software Engineering, Tongji University
Shanghai, China
cslinzhang@tongji.edu.cn

Lijun Zhang
School of Software Engineering, Tongji University
Shanghai, China
1632761@tongji.edu.cn

Xiao Liu
School of Software Engineering, Tongji University
Shanghai, China
1532787@tongji.edu.cn

Ying Shen*
School of Software Engineering, Tongji University
Shanghai, China
yingshen@tongji.edu.cn

Shaoming Zhang*
College of Surveying and Geo-informatics, Tongji University
Shanghai, China
08053@tongji.edu.cn

Shengjie Zhao
School of Software Engineering, Tongji University
Shanghai, China
shengjiezhao@tongji.edu.cn

ABSTRACT

How to restore back-lit images still remains a challenging task. State-of-the-art methods in this field are based on supervised learning and thus they are usually restricted to specific training data. In this paper, we propose a "zero-shot" scheme for back-lit image restoration, which exploits the power of deep learning, but does not rely on any prior image examples or prior training. Specifically, we train a small image-specific CNN, namely ExCNet (short for Exposure Correction Network) at test time, to estimate the "S-curve" that best fits the test back-lit image. Once the S-curve is estimated, the test image can be then restored straightforwardly. ExCNet can adapt itself to different settings per image. This makes our approach widely applicable to different shooting scenes and kinds of back-lighting conditions. Statistical studies performed on 1512 real back-lit images demonstrate that our approach can outperform the competitors by a large margin. To the best of our knowledge, our scheme is the first unsupervised CNN-based back-lit image restoration method. To make the results reproducible, the source code is available at https://cslinzhang.github.io/ExCNet/.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision problems**.

KEYWORDS

image exposure correction, image restoration, unsupervised learning

ACM Reference Format:
Lin Zhang, Lijun Zhang, Xiao Liu, Ying Shen, Shaoming Zhang, and Shengjie Zhao. 2019. Zero-Shot Restoration of Back-lit Images Using Deep Internal Learning. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19), October 21–25, 2019, Nice, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3343031.3351069

*Corresponding Author: Ying Shen and Shaoming Zhang

MM '19, October 21–25, 2019, Nice, France

ACM ISBN 978-1-4503-6889-6/19/10…$15.00
https://doi.org/10.1145/3343031.3351069

1 INTRODUCTION

When taking images, one of the most annoying types of ill illumination conditions is back-lighting. Though most modern imaging sensors can automatically adjust relevant hardware parameters (such as the aperture, the shutter speed, and the electronic gain [24]) according to the illumination conditions of the shooting targets, they still cannot get satisfied results in most back-lighting cases. One possible way to combat this problem is to use HDR (high dynamic range) imaging techniques [5]. However, HDR can only be used for image acquisition in back-lighting but cannot help restore existing poor-quality back-lit images.

In this work, we investigate the problem of back-lit image restoration and introduce a "zero-shot" scheme, namely ExCNet (short for Exposure Correction Network). By "zero-shot", we mean that ExCNet does not need prior training. ExCNet is a specially designed CNN (Convolutional Neural Networks) which can estimate the best-fitting S-curve [2, 28] for the given back-lit image directly. With its S-curve, the back-lit image can be then restored accordingly. In Fig. 1, we show 4 typical back-lit images along with the restored results by ExCNet. It can be seen that the proposed approach has a strong power for restoring poor-quality back-lit images.

1.1 Related Work

Actually, conventional image enhancement methods [8, 13, 19, 21, 30] can be explored to enhance back-lit images, but in most cases their efficacy is quite limited. There are also some heuristic methods specially designed for restoring back-lit images. In [22], Safonov proposed such an approach based on contrast stretching and alpha-blending of both brightness of the initial image and estimations of reflectance. In [25], Tsai and Yeh first detected the back-lit regions by simple thresholding the luminance channel and then they linearly stretched the detected back-lit regions. Lee *et al.* [15] refined Tsai and Yeh's idea by introducing quad-tree growing and guided filtering [11]. In [12], Im *et al.*'s approach first extracts under-exposed regions using the dark channel prior [10] and then

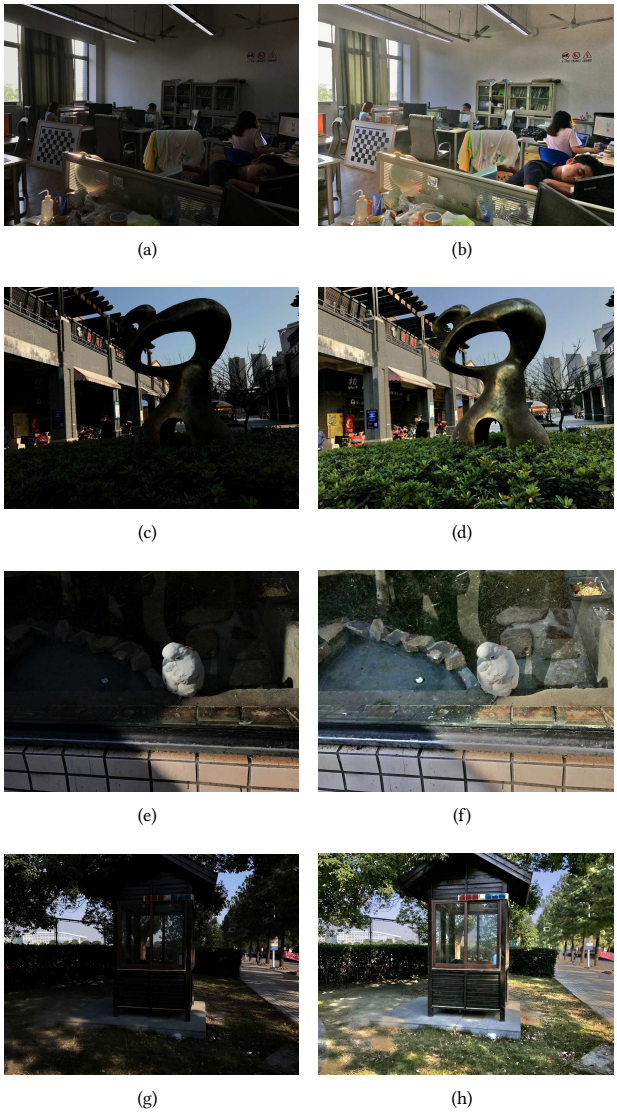

Figure 1: Back-lit image restoration results of ExCNet. Images in the first column are the original back-lit images while the ones in the second column are the enhanced results by our ExCNet.

performs spatially adaptive contrast enhancement. In [28], Yuan and Sun estimated the image specific S-curve through zone-based region-level optimal exposure evaluation. They formulated the best zone estimation as a graph labeling problem and solved it by a brute-force searching strategy. Later, the authors extended their work to correct ill-exposed video sequences [6].

Another branch of back-lit image restoration methods are based on theories of machine learning. In [4], Dale *et al.* first established a database comprising 1 million images. Given an input image to be restored, their system executes a visual search to find the closest images in the database; these images define the input's visual context, which can be further exploited to instantiate the restoration operations. In [14], Kang *et al.* constructed a database which stored the feature vectors describing training images along with vectors of enhancement parameters. Given a test image, the database was then searched for the best matching image, and the corresponding enhancement parameters were used to perform the enhancement. Following Kang *et al.*'s idea [14], Bychkovsky *et al.* [1] constructed a dataset containing 5000 example input-output image pairs that could be used to learn global tonal adjustments. In [16], Li and Wu proposed a two-phase pipeline, which performs an object-guided segmentation of back-lit and front-lit regions followed by spatially adaptive tone mapping.

1.2 Our Motivations and Contributions

We find that using machine learning tools to solve the problem of back-lit image restoration is a recent trend and also a promising direction. However, it should be noticed that existing solutions [1, 3, 4, 14, 16, 27] in this field are all based on supervised learning frameworks and thus their performance highly depends on the training dataset. In fact, for the problem of back-lit image restoration, how to collect sufficient effective training data is a challenging task itself. Consequently, these supervised learning based schemes usually perform quite well on test images satisfying the conditions on which they were trained; on the contrary, their performance deteriorates significantly once these conditions are not satisfied. This prompted us to think: Is it possible to learn an exposure correction model from the back-lit image to be restored itself? If "Yes", the learned model will be image-specific and the approach can inherently adapt itself to different settings of the unseen images. Actually, the idea of learning the restoration model solely from the test image itself has been manifested to be feasible in the field of image super-resolution [23]. In [23], Shocher *et al.* proposed a super-resolution model, which is an image-specific CNN trained on internal examples extracted solely from the low resolution test image. When dealing with real-world low resolution images, Shocher *et al.*'s method substantially outperforms its competitors.

Thus in this paper, motivated by the success of Shocher *et al.*'s work in the field of super-resolution, a "zero-shot" scheme for back-lit image restoration is proposed. The term "zero-shot" is borrowed from the domains of recognition and here is used to emphasize that our approach does not need prior image examples or prior training. Our major contributions are summarized as follows.

(1) The core of our approach is a specially designed CNN, namely ExCNet (Exposure Correction Network). Given a test back-lit image I, ExCNet could estimate the parametric "S-curve" that best fits I within limited iteration times. With its S-curve, I can be restored straightforwardly. To our knowledge, our work is the first unsupervised learning framework to correct image exposure automatically. It can be easily applied on images with various contents under different exposure levels.

(2) When designing ExCNet, a key issue is how to devise a loss function that can evaluate an image's degree of being ill-exposed. To this end, motivated by the formulation of MRF (Markov Random Field) [17], we design a block-based loss function, which tends to maximize the visibility of all blocks while keeping the relative difference between neighboring blocks. Experiments show that the designed loss function could guide ExCNet to obtain restored results

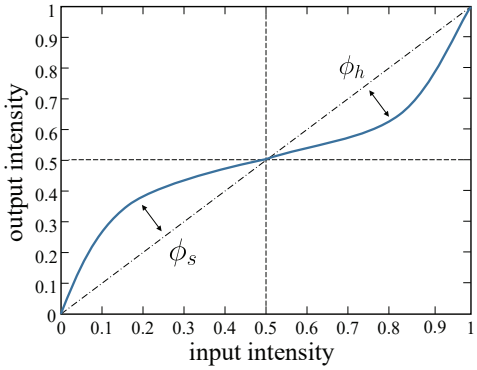

Figure 2: S-curve. ϕ_s and ϕ_h control the magnitude of S-curve adjustment in the shadow range and the highlight range, respectively.

with high visual quality. When the results of multiple competing methods are presented in front of them, people are more inclined to prefer the output of our method.

(3) Due to the CNN structure of ExCNet, our method could learn the mapping relationship between images and their best "S-curve" parameters. Thus along with the increasing of processed images, ExCNet takes less iterations to converge to the optimized curve when facing an unseen image. Besides, when handing video stream, the correction of subsequent frames could be guided by the parameters of the previous frames, which would not lead to significant flickering artifacts as [28] and have relatively low computational cost.

2 S-CURVE

Most photographers often use an S-shaped non-linear curve, namely S-curve, to manually adjust the exposure of shadow/mid-tone/highlight areas in photographs by photo editing softwares [2]. An S-curve can map input levels to desired output levels. The graph of a typical S-curve is demonstrated in Fig. 2. The horizontal axis of the graph represents the original image values while the vertical axis represents the new adjusted values. When adjusting an image, the upper-right area of the graph represents the highlights and the lower-left area represents the shadows.

As suggested in [28], the S-curve can be simply parameterized by two parameters, shadow amount ϕ_s and highlight amount ϕ_h, and accordingly it can be represented as,

$$f(x:\phi_s,\phi_h) = x + \phi_s \times f_\Delta(x) - \phi_h \times f_\Delta(1-x) \tag{1}$$

where x and $f(x:\phi_s,\phi_h)$ are the input and output luminance values in the image.[1] The incremental function $f_\Delta(t)$ is defined as $f_\Delta(t) = k_1 t \exp(-k_2 t^{k_3})$, where k_1, k_2 and k_3 are default parameters ($k_1 = 5, k_2 = 14, k_3 = 1.6$) so that $f_\Delta(t)$ will fall in $[0, 0.5]$.

The shape of the S-curve is determined by ϕ_s and ϕ_h. ϕ_s can help shift under-exposed regions into well-exposed levels, while ϕ_h can help over-exposed regions. Given a back-lit image I, to correctly restore it, we need to find the optimal parameter pair $\{\phi_s^*, \phi_h^*\}$ for I. To this end, ExCNet is proposed in this paper.

[1]The luminance of the input image is normalized to $[0, 1]$.

3 EXCNET: AN UNSUPERVISED CNN-BASED APPROACH TO ESTIMATE THE S-CURVE

Our proposed back-lit image restoration approach ExCNet is presented here in detail. ExCNet is actually a CNN designed to estimate the optimal S-curve from the luminance channel I_l of the input image I. With the estimated S-curve, I can be restored accordingly. The structure of ExCNet is depicted in Fig. 3. To better understand it, each training iteration of ExCNet can be conceptually considered as having two stages, adjusting I_l using the intermediate S-curve and deriving the loss. Details of these two stages will be introduced in Sect. 3.1 and Sect. 3.2, respectively. Implementation details will be given in Sect. 3.3.

3.1 Adjusting I_l Using the Intermediate S-curve

For each training iteration, at this stage, I_l first goes through an internal CNN. This CNN comprises a stack of convolutional layers, followed by several fully connected layers. Its last fully connected layer has two outputs, which are regarded as the intermediate estimations of the shadow amount and the highlight amount and are denoted by $\widehat{\phi_s}$ and $\widehat{\phi_l}$, respectively. With $\widehat{\phi_s}$ and $\widehat{\phi_l}$, the intermediate S-curve is instantiated as $f\left(x;\widehat{\phi_s},\widehat{\phi_h}\right)$ (See Eq. 1). Then, I_l is corrected using $f\left(x;\widehat{\phi_s},\widehat{\phi_h}\right)$ and the result is denoted by I_l^c.

3.2 Deriving the Loss

To update the weights of ExCNet, we need to have a loss function that can evaluate the degree of being ill-exposed of the intermediate restoration result I_l^c. If we have such an ideal loss function, by minimizing it iteratively, the optimal parameters $\left\{\phi_s^*, \phi_h^*\right\}$ of the S-curve for the input image I can be obtained.

In this paper, we formulate the loss function (taking I_l^c as the input) of ExCNet as a block-based energy minimization problem. Compared with the pixel-based formulation, the block-based one has the advantage that regions can better represent the visibility of contents and measure the relative difference of neighboring regions. When designing the energy function, we should keep in mind that the back-lit image restoration operation needs to maximize the visibility of each block while keeping the relative difference between two neighboring regions. Motivated by the formulation of MRF [17], we define the loss function of ExCNet as,

$$\mathcal{L} = \sum_i (E_i + \lambda \sum_{j\in\Omega(i)} E_{ij}) \tag{2}$$

where E_i is the unary data term, E_{ij} is referred to as the pairwise term, and λ is a predefined constant. For a given block i, by minimizing E_i, its visibility is anticipated to increase. E_{ij} represents the change of the relative difference between neighboring blocks i and j before and after the restoration operation. $j \in \Omega(i)$ means that block j belongs to the set of neighboring blocks of block i. In this paper, $\Omega(i)$ is the set of 4-neighboring blocks of block i. By minimizing $\mathcal{L}$, we expect that the visibility of image blocks can increase and at the same time the relative difference between neighboring blocks can be kept as much as possible.

The unary term E_i is defined as,

$$E_i = sign\left(l_i^c - 0.5\right) \cdot \left(l_i^c - l_i\right) \tag{3}$$

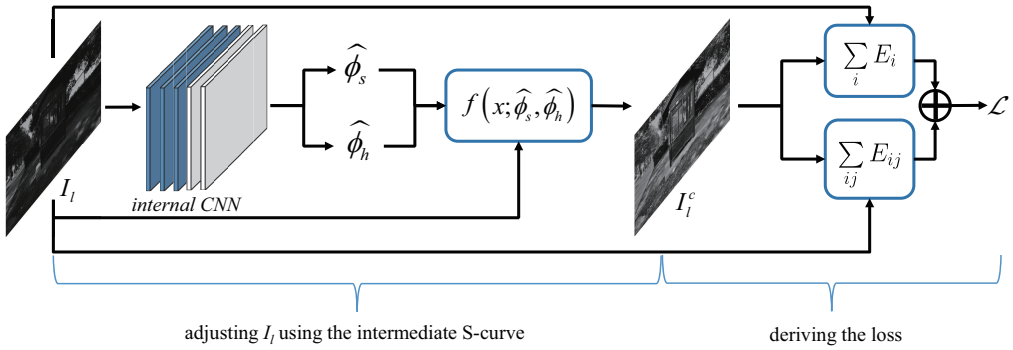

Figure 3: The structure of ExCNet. Each training iteration of ExCNet can be conceptually considered as having two stages, adjusting I_l using the intermediate S-curve and deriving the loss.

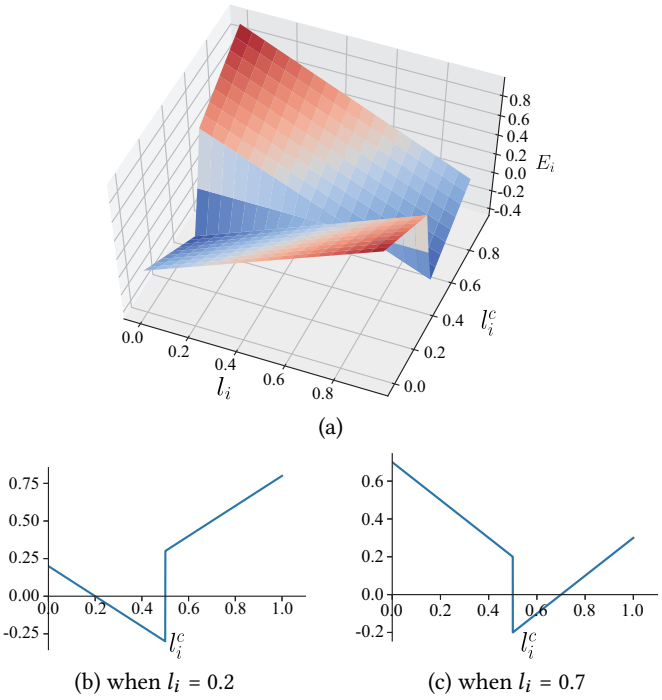

Figure 4: (a) shows the value surface of E_i as a function of l_i and l_i^c. The two examples shown in (b) and (c) demonstrate how E_i changes with l_i^c when l_i is fixed. In (b), l_i is fixed as 0.2 and in (c), l_i is fixed as 0.7.

where l_i and l_i^c are the average luminance of block i in I_l and I_l^c, respectively. Following the assumption widely adopted in HDR [20] and exposure fusion [18] field, E_i measures the visibility of each block by the distance between the block average intensity and the mid gray (0.5). Fig. 4(a) shows the value surface of E_i as a function of l_i and l_i^c. The two examples shown in Figs. 4(b) and 4(c) demonstrate how E_i changes with l_i^c when l_i is fixed. In Fig. 4(b), l_i is fixed as 0.2 (smaller than 0.5) and in Fig. 4(c), l_i is fixed as 0.7 (greater than 0.5). It can be easily verified that the unary term

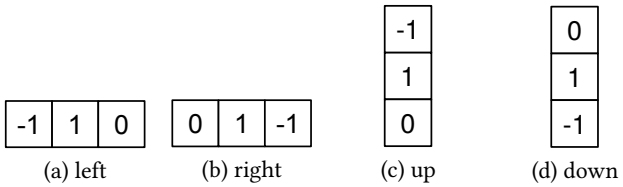

Figure 5: 4 non-trainable convolution kernels, each for computing the block difference in one direction.

E_i encourages l_i^c as close to 0.5 as possible so that both the under- and over-exposed regions can be adjusted to well-exposed levels. In addition, E_i encourages l_i^c to inherit the value relationship between l_i and 0.5. That is to say, if l_i is greater (smaller) than 0.5, l_i^c will also tend to be greater (smaller) than 0.5. Such a property of E_i can guarantee that lighter/darker regions in I_l remain lighter/darker in I_l^c.

The pairwise term E_{ij} is defined as,

$$E_{ij} = \left(\left(l_j^c - l_i^c\right) - \left(l_j - l_i\right)\right)^2 \tag{4}$$

where l_j (l_j^c) and l_i (l_i^c) are the average luminance of block j and block i, respectively, in I_l (I_l^c). This term is designed to keep the relative difference between neighboring regions.

The combination of the unary term and the pairwise term can encourage exposure to get close to 0.5 but will stop at some intensity by the pairwise term so that the relative difference between neighboring regions can be kept.

3.3 Implementation Details

ExCNet is implemented as a CNN, whose detailed configurations are summarized in Table. 1. ExCNet takes the normalized luminance map I_l with the fixed size 128×128 as the input.

One key issue in implementation is how to compute the block-based loss function with plain convolution operations. A 4×4 average pooling is applied to I_l (I_l^c) and the result is denoted by I_{lb} (I_{lb}^c). In this way, the value of a "pixel" in I_{lb} (I_{lb}^c) is actually the average luminance of a particular 4×4 block in I_l (I_l^c). The unary term E_i (Eq. 3) can be easily computed with I_{lb} and I_{lb}^c.

Table 1: Detailed configurations of ExCNet.

Stage	Type	Input	Output	Output Dimension
Input	Input	-	I_l	(128, 128, 1)
Adjusting I_l using the intermediate S-curve	Conv + Pooling (kernel size: 3×3)	I_l	$Conv_1$	(64,64,64)
		$Conv_1$	$Conv_2$	(32, 32, 128)
		$Conv_2$	$Conv_3$	(16, 16, 256)
		$Conv_3$	$Conv_4$	(8, 8, 512)
		$Conv_4$	$Conv_5$	(4, 4, 512)
	Fully connected	$Conv_5$	FC_1	(512, 1)
		FC_1	FC_2	(256, 1)
		FC_2	FC_3	(128, 1)
		FC_3	$\widehat{\phi_s}, \widehat{\phi_h}$	-
	$f(x : \widehat{\phi_s}, \widehat{\phi_h})$ (Eq. 1)	$I_l, \widehat{\phi_s}, \widehat{\phi_h}$	I_l^c	(128, 128, 1)
Deriving the loss	Average pooling	I_l, I_l^c	I_{lb}, I_{lb}^c	(32, 32, 1)
	Conv (4 non-trainable kernels, shown in Fig. 5)	I_{lb}, I_{lb}^c	$\{D_{lb}^r\}, \{D_{lb}^{cr}\}$	(32, 32, 1)
	E_{data} (Eq. 3)	I_{lb}, I_{lb}^c	E_{data}	-
	E_{smooth} (Eq. 4)	$\{D_{lb}^r\}, \{D_{lb}^{cr}\}$	E_{smooth}	-
	Sum (Eq. 2)	E_{data}, E_{smooth}	$\mathcal{L}$	-

To get the differences between one block and its 4-neighboring blocks in I_l (I_l^c), we can convolve I_{lb} (I_{lb}^c) with 4 non-trainable one-dimensional kernels as shown in Fig. 5 and the results are denoted by $\left\{D_{lb}^r : |r \in \{left, right, up, down\}\right\}$ ($\left\{D_{lb}^{cr}\right\}$). The pairwise term E_{ij}, which depends on the differences between adjacent blocks, can be computed straightforwardly from $\left\{D_{lb}^r\right\}$ and $\left\{D_{lb}^{cr}\right\}$.

ExCNet can be considered as an unsupervised learning approach. That is to say, after processing some images, ExCNet can progressively learn a capability to capture the image's tonality information. When dealing with forthcoming images, such a pre-trained ExCNet could converge much faster than its counterpart which is randomly initialized. Thus, in our implementation, we prepared 50 back-lit images and used ExCNet to estimate their optimal S-curves. The final weights were saved and used afterwards for initializing ExCNet when coping with new images.

4 BACK-LIT IMAGE RESTORATION USING EXCNET

The overall pipeline of the proposed back-lit image restoration scheme is presented in this section. Suppose that I is the given back-lit image to be restored. Denote by $\hat{I}$ the restoration result. The restoration pipeline comprises three major steps, S-curve estimation, luminance channel restoration, and chrominance adjustment. I is first transformed from the RGB space to the YIQ space as (I_Y, I_I, I_Q), where I_Y, I_I, I_Q are the three YIQ channels. After that, the data range of I_Y is normalized to [0, 1]. To meet ExCNet's requirement of the input size, I_Y is then resized to 128×128 and the result is regarded as I_l. Afterwards, the optimal S-curve parameterized by $\{\phi_s^*, \phi_h^*\}$ is estimated by feeding I_l to ExCNet. With I_Y and $\{\phi_s^*, \phi_h^*\}$, the luminance channel $\hat{I}_Y$ of the final restoration result $\hat{I}$ can be derived using Eq. 1. The two chrominance channels I_I and I_Q are adjusted accordingly as,

$$\begin{cases} \hat{I}_I(\mathbf{x}) = \frac{\hat{I}_Y(\mathbf{x})}{I_Y(\mathbf{x})} \times I_I(\mathbf{x}) \\ \hat{I}_Q(\mathbf{x}) = \frac{\hat{I}_Y(\mathbf{x})}{I_Y(\mathbf{x})} \times I_Q(\mathbf{x}) \end{cases} \tag{5}$$

where $\hat{I}_I$ and $\hat{I}_Q$ are the two chrominance channels of the final restoration result $\hat{I}$, and $\mathbf{x}$ is the pixel's position. Finally, convert $\hat{I}$ from the YIQ space to the RGB space.

It is worth mentioning that we also take detail preserving into consideration in our implementation. We firstly separate each input image into a base layer and a detail layer using the guided filter [11] and then use the estimated S-curve to adjust the base layer. Finally, we adopt Weber contrast [26] to fuse the detail layer and the adjusted base layer.

Our hardware platform is a workstation with a 3.0GHZ Intel Core i7-5960X CPU and an Nvidia Titan X GPU card. The method is implemented with Python and TensorFlow. It costs our method about 1.0s to process one 4032×3024 image.

5 EXPERIMENTAL RESULTS

5.1 Dataset and the Compared Methods

Experiments were performed on 1,512 real-world images taken from $IE_{ps}D$ [29], which was established for studying the problem of exposure level assessment. These images vary on scenes and lighting conditions. We invited three subjects to roughly partition these images into three groups, around 500 images for each group, according to their exposure conditions. Three groups are "severely

Figure 6: Results of pairwise comparison user study. The pairwise comparison was conducted between the result of ExCNet and the input or the result of one of the eight compared methods. Input images were from "Group A", "Group B", "Group C", and "All Groups", respectively. Each color bar denotes the average percentage of the image version favored (with I-shaped standard deviation bars). (Best viewed on screen.)

ill-exposed" (Group A), "slightly ill-exposed" (Group B), and "well-exposed" (Group C).

The proposed method ExCNet was compared with eight state-of-the-art or representative automatic exposure correction algorithms, including 1) HE [21], 2) CLAHE [30], 3) Retinex [13], 4) Google Picasa's *Auto-contrast* [9], 5) edge-preserving decomposition (wls-filter) [8], 6) local Laplacian filtering (lapfilter) [19], 7) Yuan and Sun's method [28], and 8) Li and Wu's method [16].

5.2 Pairwise Comparison User Studies

Ten volunteers (6 males and 4 females) were invited to perform pairwise comparison between our result and the input or the result of one of the eight compared methods. For each pairwise comparison, the subject had three options: "left better", "right better", or "no preference".

The results of the user study are summarized in graphs shown in Fig. 6. Each color bar is the averaged percentage of the image version favored over all 10 subjects (I-shape error bar denotes the standard deviation). From results on "All Groups" (without distinguishing the photos from different groups), it is obvious that the participants overwhelmingly selected our results over the input or the outputs of the compared methods. The participants showed a strong bias in preference towards our correction results when compared to CLAHE [30], edge-preserving decomposition (wlsfilter) [8], and local Laplacian filtering (lapfilter) [19]. Moreover, ExCNet gained

60%∼70% of the favor compared with the input or the results by HE [21], Retinex [13], Picasa [9], Yuan and Sun's method [28], and Li and Wu's method [16]. The results from "Group A" and "Group B" indicate that ExCNet performed significantly better for restoring both severely ill-exposed and slightly ill-exposed images than its competitors. When dealing with properly-exposed images in "Group C", ExCNet also performed well. For about 84% images in "Group C", ExCNet did not make the results worse compared with the inputs. These results imply that ExCNet can work consistently well when processing images with different lighting conditions. By contrast, some approaches (such as HE [21], Yuan and Sun's method [28], and Li and Wu's method [16]) may achieve acceptable results when processing ill-exposed images but they tend to deteriorate properly-exposed ones. This phenomenon can be clearly observed through examples shown in Fig. 7.

5.3 Visual Quality Comparisons

In order to facilitate readers to visually compare the results of different back-lit image restoration approaches, we show several examples in Fig. 7. Fig. 7(a) (b or c) shows one typical image selected from Group A (B or C) along with the results obtained by mainstream schemes in this field. From the results shown in Fig. 7, we can have the following findings.

It is evident that the results obtained by the two conventional methods for image enhancement, HE [21] and Retinex [13], suffer from severe color deviation. When processing severely or slightly ill-exposed images, neither Picasa [9] nor lapfilter [19] has a satisfied capability for increasing the exposure levels of under-exposed regions. It seems that Yuan and Sun's method [28] and Li and Wu's method [16] can work well for restoring ill-exposed images. However, both of them tend to destroy properly-exposed inputs and such a fact can be observed through examples shown in Fig. 7(c). In Fig. 7(c), the input image is properly-exposed and is of high visual quality. Yuan and Sun's method [28] decreases its contrast and Li and Wu's method [16] introduces annoying halo effects. It implies that these two approaches cannot be routinely applied to imaging systems where whether the acquired images are properly-exposed or not cannot be guaranteed. By contrast, the proposed method ExCNet can produce pleasing results for all cases, indicating that it is more robust and can be applied to a wider range of applications. A demo video is provided in the supplementary material.

5.4 Objective Evaluation

The performance of back-lit image restoration methods was also evaluated with two objective metrics, CDIQA (contrast-distorted image quality assessment) [7] and LOD (luminance ordinal distortion) [16]. CDIQA is a no-reference quality metric for contrast-distorted images, which can be considered as a metric for richness of image details. Higher CDIQA value roughly corresponds to higher contrast. LOD is defined as,

$$LOD = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\|\hat{\mathbf{v}}_i - \mathbf{v}_i\|^2} \tag{6}$$

where N is the number of sliding windows, $\mathbf{v}_i$ and $\hat{\mathbf{v}}_i$ are the pixels' luminance ordinal vectors of the i-th window in the input and corrected images, respectively. Ideally, if the restoration approach does not violate the order statistics of pixel values of the input image, the associated LOD measure would be zero. Unwanted artifacts introduced by the restoration algorithm, such as contours, halos and rings, could lead to changes in order statistics. Thus, LOD can measure the severity of artifacts caused by restoration.

Table 2: Objective evaluation results.

Methods	CDIQA	LOD
HE [21]	2.8757	4.4820
CLAHE [30]	3.0602	3.5214
Retinex [13]	3.2021	3.9602
Picasa [9]	3.0667	**2.2694**
wlsfilter [8]	2.7608	3.7365
lapfilter [19]	2.7790	5.0398
Yuan and Sun [28]	2.9451	4.6261
Li and Wu [16]	3.2494	4.9643
ExCNet	**3.2616**	**2.8030**

The results over 1,512 test images are reported in Table. 2. From Table. 2, it can be seen that ExCNet achieves the highest CDIQA value, demonstrating its superiority in restoring image details over the other methods. With respect to LOD, ExCNet and Picasa [9] are comparable and both of them could achieve quite low LOD values than the other competitors. Such a result demonstrates that compared with ExCNet and Picasa the other methods are prone to annoying over-enhancement artifacts. On the other hand, it is not surprising that Picasa could achieve the least LOD value since Picasa actually has a weak capability to restore ill-exposed images and thus the order statistics could be kept well. Overall, it can be seen that among the evaluated schemes ExCNet is the most competent one for restoring back-lit images without introducing objectionable artifacts.

6 CONCLUSION

Back-lit image restoration is of practical importance and has not been well resolved. In this paper, a "zero-shot" back-lit image restoration scheme is proposed, which exploits the power of deep learning, without relying on any external examples or prior training. This is achieved via a small image-specific CNN, namely ExCNet, which guides the restoration progress by minimizing a block-based loss function defined on the intermediate restoration result. The proposed scheme is concise yet powerful. It is quite robust and thus can yield pleasing results under different kinds of illumination conditions. Subjective and objective evaluations were conducted comprehensively to corroborate the superiority of the proposed approach over the other mainstream competitors. To our knowledge, it is the first unsupervised CNN-based back-lit image restoration method.

ACKNOWLEDGMENTS

This work was supported in part by the Natural Science Foundation of China under grant no. 61672380, in part by the Natural Science Foundation of Shanghai under grant no. 19ZR1461300, and in part by the Science and Technology Commission of Shanghai under grant no. 17DZ1100202.

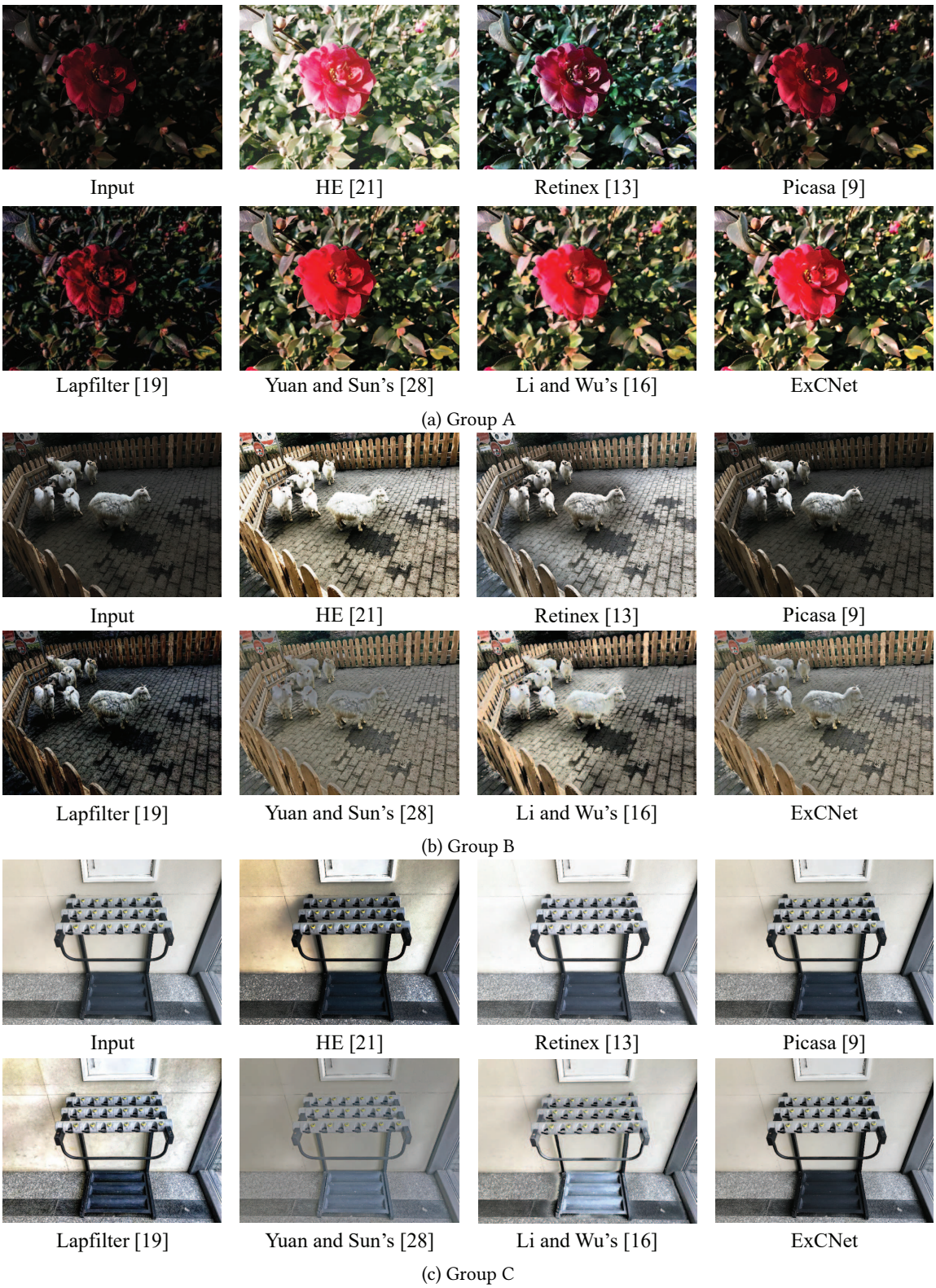

Figure 7: Visual quality comparison of results obtained by different back-lit image restoration approaches. Input images in (a), (b), and (c) are selected from "Group A", "Group B", and "Group C", respectively. (Best viewed on screen.)

REFERENCES

[1] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. 2011. Learning photographic global tonal adjustment with a database of input/output image pairs. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*. IEEE, Colorado Springs, CO, USA, 97–104. https://doi.org/10.1109/CVPR.2011.5995413

[2] Photoshop CC. 2017. Adjust color and tonality with curves. https://helpx.adobe.com/photoshop/using/curves-adjustment.html#id_82503

[3] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. 2018. Learning to See in the Dark. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '18)*. IEEE, Salt Lake City, UT, USA, 3291–3300. https://doi.org/10.1109/CVPR.2018.00347

[4] Kevin Dale, Micah K. Johnson, Kalyan Sunkavalli, Wojciech Matusik, and Hanspeter Pfister. 2009. Image restoration using online photo collections. In *2009 IEEE 12th International Conference on Computer Vision (ICCV '09)*. IEEE, Kyoto, Japan, 2217–2224. https://doi.org/10.1109/ICCV.2009.5459473

[5] Paul E. Debevec and Jitendra Malik. 2008. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes (SIGGRAPH '08)*. ACM, New York, NY, USA, 369–378. https://doi.org/10.1145/1401132.1401174

[6] Xuan Dong, Lu Yuan, Weixin Li, and Alan L. Yuille. 2015. Temporally consistent region-based video exposure correction. In *2015 IEEE International Conference on Multimedia and Expo (ICME '15)*. IEEE, Turin, Italy, 1–6. https://doi.org/10.1109/ICME.2015.7177382

[7] Yuming Fang, Kede Ma, Zhou Wang, Weisi Lin, Zhijun Fang, and Guangtao Zhai. 2015. No-reference quality assessment of contrast-distorted images based on natural scene statistics. *IEEE Signal Processing Letters* 22, 7 (July 2015), 838–842. https://doi.org/10.1109/LSP.2014.2372333

[8] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph.* 27, 3, Article 67 (Aug. 2008). https://doi.org/10.1145/1399504.1360666

[9] Google. 2013. Picasa. http://picasa.google.com

[10] Kaiming He, Jian Sun, and Xiaoou Tang. 2011. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 12 (Dec. 2011), 2341–2353. https://doi.org/10.1109/TPAMI.2010.168

[11] Kaiming He, Jian Sun, and Xiaoou Tang. 2013. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 6 (June 2013), 1397–1409. https://doi.org/10.1109/TPAMI.2012.213

[12] Jaehyun Im, Inhye Yoon, Monson H. Hayes, and Joonki Paik. 2013. Dark channel prior-based spatially adaptive contrast enhancement for back lighting compensation. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '13)*. IEEE, Vancouver, BC, Canada, 2464–2468. https://doi.org/10.1109/ICASSP.2013.6638098

[13] Daniel J. Jobson, Ziaur Rahman, and Glenn A. Woodell. 1997. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Processing* 6, 7 (July 1997), 965–976. https://doi.org/10.1109/83.597272

[14] Sing Bing Kang, Ashish Kapoor, and Dani Lischinski. 2010. Personalization of image enhancement. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '10)*. IEEE, San Francisco, CA, USA, 1799–1806. https://doi.org/10.1109/CVPR.2010.5539850

[15] Seungwon Lee, Nahyun Kim, and Joonki Paik. 2015. Adaptively partitioned block-based contrast enhancement and its application to low light-level video surveilance. *SpringerPlus* 4, 1, Article 431 (Aug. 2015). https://doi.org/10.1186/s40064-015-1226-x

[16] Zhenhao Li and Xiaolin Wu. 2018. Learning-based restoration of backlit images. *IEEE Trans. Image Processing* 27, 2 (Feb. 2018), 976–986. https://doi.org/10.1109/TIP.2017.2771142

[17] Ziwei Liu, Xiaoxiao Li, Ping Luo, Chen Change Loy, and Xiaoou Tang. 2018. Deep learning Markov random field for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 8 (Aug. 2018), 1814–1828. https://doi.org/10.1109/TPAMI.2017.2737535

[18] Tom Mertens, Jan Kautz, and Frank Van Reeth. 2007. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Applications (PG '07)*. IEEE, Maui, HI, USA, 382–390. https://doi.org/10.1109/PG.2007.17

[19] Sylvain Paris, Samuel W. Hasinoff, and Jan Kautz. 2015. Local Laplacian filters: Edge-aware image processing with a Laplacian pyramid. *Commun. ACM* 58, 3 (March 2015), 81–91. https://doi.org/10.1145/2723694

[20] Mark A. Robertson, Sean Borman, and Robert L. Stevenson. 2003. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging* 12, 2 (April 2003), 219–228. https://doi.org/10.1117/1.1557695

[21] John C Russ. 2011. *The Image Processing Handbook*. CRC Press, Boca Raton.

[22] Ilia V. Safonov. 2006. Automatic correction of amateur photos damaged by backlighting. In *16th International Conference on Computer Graphics and Vision (GraphiCon '06)*. GRAPHICON, 80–89.

[23] Assaf Shocher, Nadav Cohen, and Michal Irani. 2018. "Zero-shot" super-resolution using deep internal learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '18)*. IEEE, Salt Lake City, UT, USA, 3118–3126. https://doi.org/10.1109/CVPR.2018.00329

[24] Juan Torres and José Manuel Menéndez. 2015. Optimal camera exposure for video surveillance systems by predictive control of shutter speed, aperture, and gain. In *Real-Time Image and Video Processing 2015*. International Society for Optics and Photonics, San Francisco, CA, USA, 94000S:1–14. https://doi.org/10.1117/12.2083182

[25] Chun Ming Tsai and Zong Mu Yeh. 2010. Contrast compensation by fuzzy classification and image illumination analysis for back-lit and front-lit color face images. *IEEE Trans. Consum. Electron.* 56, 3 (Aug. 2010), 1570–1578. https://doi.org/10.1109/TCE.2010.5606299

[26] Paul Whittle. 1994. The psychophysics of contrast brightness. *A. L. Gilchrist (Ed.), Brightness, lightness, and transparenc* (1994), 35–110.

[27] Xin Yang, Ke Xu, Yibing Song, Qiang Zhang, Xiaopeng Wei, and Rynson WH. Lau. 2018. Image Correction via Deep Reciprocating HDR Transformation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '18)*. IEEE, Salt Lake City, UT, USA, 1798–1807. https://doi.org/10.1109/CVPR.2018.00193

[28] Lu Yuan and Jian Sun. 2012. Automatic exposure correction of consumer photographs. In *12th European Conference on Computer Vision (ECCV '12)*. Springer, Berlin, Heidelberg, 771–785. https://doi.org/10.1007/978-3-642-33765-9_55

[29] Lijun Zhang, Lin Zhang, Xiao Liu, Ying Shen, and Dongqing Wang. 2018. Image exposure assessment: A benchmark and a deep convolutional neural networks based model. In *2018 IEEE International Conference on Multimedia and Expo (ICME '18)*. IEEE, San Diego, CA, USA, 1–6. https://doi.org/10.1109/ICME.2018.8486569

[30] Karel Zuiderveld. 1994. Contrast limited adaptive histogram equalization. In *Graphics Gems IV*. Academic Press Professional, Inc., San Diego, CA, USA, 474–485. http://dl.acm.org/citation.cfm?id=180895.180940