Pattern Recognition Letters 104 (2018) 15–22

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

LightenNet: A Convolutional Neural Network for weakly illuminated image enhancement

Chongyi Li[a,b], Jichang Guo[a,*], Fatih Porikli[b], Yanwei Pang[a]

[a] School of Electrical and Information Engineering, Tianjin University, Weijing Road 92, Tianjin 300300, China
[b] Research School of Engineering, College of Engineering and Computer Science, Australian National University, Canberra, ACT 0200, Australia

ARTICLE INFO

Article history:
Received 8 July 2017
Available online 16 January 2018

MSC:
41A05
41A10
65D05
65D17

Keywords:
Low light image enhancement
Weak illumination image enhancement
Image degradation
CNNs

ABSTRACT

Weak illumination or low light image enhancement as pre-processing is needed in many computer vision tasks. Existing methods show limitations when they are used to enhance weakly illuminated images, especially for the images captured under diverse illumination circumstances. In this letter, we propose a trainable Convolutional Neural Network (CNN) for weakly illuminated image enhancement, namely LightenNet, which takes a weakly illuminated image as input and outputs its illumination map that is subsequently used to obtain the enhanced image based on Retinex model. The proposed method produces visually pleasing results without over or under-enhanced regions. Qualitative and quantitative comparisons are conducted to evaluate the performance of the proposed method. The experimental results demonstrate that the proposed method achieves superior performance than existing methods. Additionally, we propose a new weakly illuminated image synthesis approach, which can be use as a guide for weakly illuminated image enhancement networks training and full-reference image quality assessment.

1. Introduction

High quality images and videos are needed in both computer vision applications (*e.g.*, object detection and tracking) and consumer electronics (*e.g.*, hand-hold devices). However, images and videos captured under weak illumination conditions often suffer from noticeable degradation of visibility, brightness, and contrast. Accordingly, weak illumination image enhancement as pre-processing is of significance and also desired.

Despite numerous methods [4–7,9,11,19–21,24,28] have been developed for degraded image and video enhancement, there are still some issues that remain to be resolved for weak illumination image enhancement. Therefore, we propose a new method to enhance images captured under poorly-lit circumstances. First, we build our method on the Retinex model [18] which can be used to enhance image via estimating its illumination map. Thus, the key to achieve enhanced image is to estimate an accurate illumination map. Then, we formulate a compact and efficient CNN, namely LightenNet, which takes a weakly illuminated image as input and outputs its illumination map that is subsequently used to obtain the enhanced image based on Retinex model. Both qualitatively and quantitatively experimental results show that our method generates accurate illumination maps and achieves more natural-looking results and better details than the state-of-the-art methods. Additionally, our method generalizes well to the images captured under varying illumination conditions. Here, varying illumination conditions represent non-uniform illumination, back lighting, extremely weak illumination intensity, well-lit, and so forth. The main contributions of this letter are summarized as follows:

- We propose a trainable CNN for weak illumination image enhancement. Different from several existing CNN-based image enhancement methods which directly estimate the enhanced or restored image, our LightenNet learns to predict the mapping relations between weakly illuminated image and the corresponding illumination map. Thus, LightenNet is easy to be trained. It just takes 1 h to optimize our LightenNet.
- Based on Retinex model, we propose a new method for weakly illuminated image synthesis, which can be use as a guide for subsequent network training and full-reference image quality assessment.
- Compared to existing weak illumination image enhancement methods, our CNN-based method achieves the state-of-the-art performance on both synthetic and real weakly illuminated images.

The remainder of the letter is organized as follows. Section 2 presents a brief overview of the related work. Section 3 first reviews Retinex model, and then introduces

* Corresponding author.
E-mail address: jcguo@tju.edu.cn (J. Guo).

https://doi.org/10.1016/j.patrec.2018.01.010
0167-8655/

the proposed LightenNet model. Section 4 shows experimental results. Section 5 concludes this letter.

2. Related work

Over the past few decades, many methods have been proposed to enhance degraded images. However, fewer methods can generate ideal results when they are applied to weakly illuminated images. For example, histogram-based methods [23] usually increase the dynamic range of the gray values, which leads to suboptimum enhancement performance for image details. For another example, physical model-based methods [19,28] usually produce unnatural and unrealistic results since some priors or assumptions do not always hold for varying illumination conditions.

In contrast to traditional image enhancement methods, there are fewer methods that enhance weakly illuminated images. Several existing enhancement methods [4,19,28] are based on the observation that the inverted low-light images intuitively look like haze images. Such a method first inverts an input low-light image, and then employs an image dehazing algorithm on the inverted image, finally achieves the enhanced image by inverting the dehazed image. Dehazing-like methods can enhance the visual quality of low-light images to some extent, however, these methods lack of cogent physical explanation and tend to produce unrealistic results. Recently, Fotiadou et al. [5] proposed a novel method to enhance low-light images based on the framework of Sparse Representations. Fotiadou et al. used two dictionaries (*i.e.*, *night dictionary* and *day dictionary*) to transform the Sparse Representation of low-light image patches to the corresponding enhanced image patches. The enhanced results significantly relied on the accuracy of the learned dictionaries. Fu et al. [7] employed fusion-based method for weakly illuminated image enhancement, which fused luminance-improved and contrast-enhanced versions of input by two designed weights. Besides, multi-scale fusion scheme was applied to reduce the amplified artifacts. The enhanced results are characteristic with improved brightness, contrast, and details. However, like other fusion-based image enhancement methods, such a method tends to produce over-enhanced, over-saturated, and unrealistic results due to ignoring the physical properties of weak illumination image degradation. Lore et al. [20] proposed a deep learning-based method to adaptively enhance and denoise images captured under low-light environments, namely LLNet. Lore et al. directly employed an existing deep neural network architecture (*i.e.*, stack sparse denoising autoencoder) to build the relations between the low-light images and the corresponding enhanced and denoised images. The experimental results demonstrated that deep learning-based method is suitable for low light image enhancement. Guo et al. [11] proposed a simple low-light image enhancement method, namely LIME. This method first estimated the illumination of each pixel in the low-light image, then refined the initial illumination map by a structure prior, finally the enhanced image was achieved based on Retinex model using the estimated illumination map. Besides, in order to reduce the amplified noise, an existing image denoising algorithm was used as post-processing in the LIME method.

3. The proposed method

In this section, we begin by describing Retinex model where we built our method on, and then present the layer design of the proposed LightenNet and introduce how to enhance weakly illuminated images by the predicted illumination map, last introduce the implementation details of our LightenNet.

3.1. Retinex model

Following Retinex model proposed by Land [18] which describes the lightness and color perception of human vision, we explain the formation of the weakly illuminated image as follows

$$I(x) = R(x) \cdot L(x). \tag{1}$$

where x denotes a pixel, $I(x)$ is the observed image, $R(x)$ is the reflectance of the image, and $L(x)$ is the illumination. The dot "$\cdot$" means pixel-wise multiplication. The physical meaning of Eq. (1) can be simply described as that the observed image (*i.e.*, weakly illuminated image) can be decomposed into the product of the reflectance of image $R(x)$ (*i.e.*, desired illuminated image) and the illumination map $L(x)$ (*i.e.*, dark veiling which degrades input image). Thus, weak illumination image enhancement means removing weak illumination from input image. In this letter, our goal is to achieve the reflectance $R(x)$ from the observed image $I(x)$ by predicting its illumination map $L(x)$.

3.2. LightenNet

The proposed LightenNet is inspired by the success of deep learning in low-level vision tasks [1–3]. The purpose of LightenNet is to learn a mapping, which takes a weakly illuminated image as input and outputs its illumination map that is subsequently used to obtain the enhanced image based on Retinex model. The architecture of LightenNet is shown in Fig. 1. In the LightenNet architecture, input is the weakly illuminated image and the output is the corresponding illumination map. Similar with Dong et al. [3], LightenNet contains four convolutional layers with specific tasks. Observing the feature maps in Fig. 1, different convolutional layers have different effects on final illumination map. For example, the first two layers focus on the high light regions and the third layer focuses on low light regions while the last layer is to reconstruct the illumination map. The specific operation form of the four convolutional layers is described as follows.

Patch extraction and representation: To learn the relations between weakly illuminated image and its illumination map, we first extract overlapping image patches from weakly illuminated image and represents each image patch using a high-dimensional vector by n_1 filters, and can be expressed as

$$F_1(P) = \max(0, W_1 * P + B_1), \tag{2}$$

where P is an input image patch with size $n \times n$, W_1 and B_1 are the weights and biases of filters. The size of W_1 is $f_1 \times f_1 \times n_1$, where f_1 is the spatial support of a filter, and n_1 is the number of filters. B_1 is an n_1-dimensional vector, whose each element is associated with a filter. "$*$" represents the convolution operation. $max(0, x)$ is a Rectified Linear Unit (ReLU) [17], which is applied to accelerate training convergence and improve performance.

Feature enhancement: Inspired by the feature enhancement layer used in compression artifact reduction [3], we employ a feature enhancement layer to map the "noisy" feature to a relatively "cleaner" feature space since the weakly illuminated images usually suffer from the effects of noise. It can be expressed as

$$F_2(P) = \max(0, W_2 * F_1(P) + B_2), \tag{3}$$

where W_2 contains n_2 filters of size $f_2 \times f_2 \times n_1$, and B_2 is an n_2-dimensional vector.

Non-linear mapping: The next step is to map each high-dimensional vector onto another high-dimensional vector, namely transforming $F_2(P)$ into $F_3(P)$:

$$F_3(P) = \max(0, W_3 * F_2(P) + B_3), \tag{4}$$

where W_3 contains n_3 filters of size $f_3 \times f_3 \times n_2$, and B_3 is an n_3-dimensional vector.

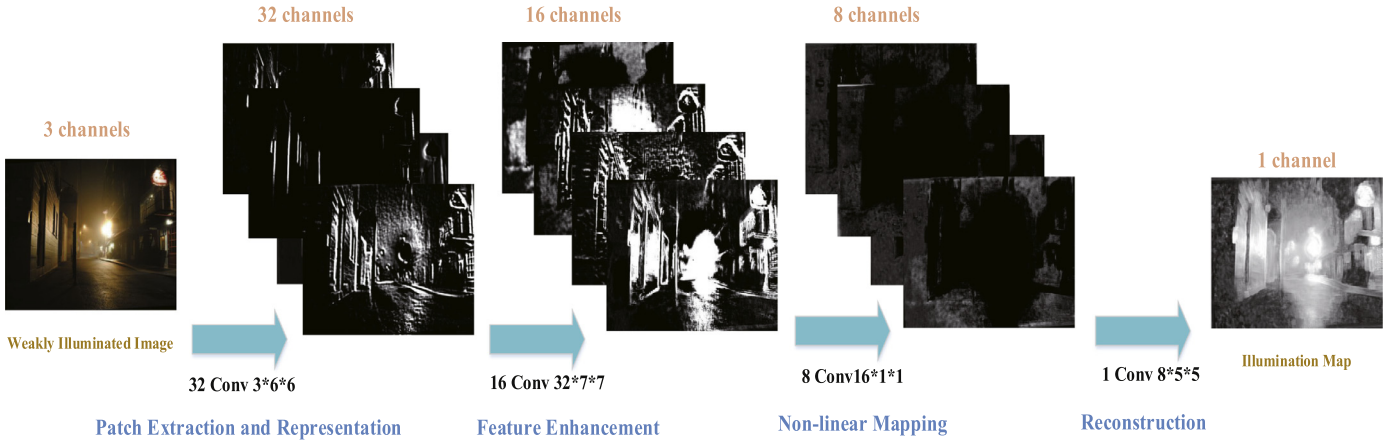


Fig. 1. The architecture of LightenNet. LightenNet consists of 4 convolution layers, *i.e.*, patch extraction and representation, feature enhancement, non-linear mapping, and reconstruction.

Reconstruction: Lastly, a convolutional layer is designed to aggregate the patch-wise representations to generate the learned illumination map. $F_3(P)$ is transformed into $F_4(P)$, and can be expressed as

$$F_4(P) = W_4 * F_3(P) + B_4. \tag{5}$$

where W_4 contains n_4 filters of size $f_4 \times f_4 \times n_3$, and B_4 is an n_4-dimensional vector.

These unknown network parameters $\Theta = \{W_1, W_2, W_3, W_4, B_1, B_2, B_3, B_4\}$ are achieved by minimizing Mean Squared Error (MSE) loss function by supervised learning manner. MSE loss function is expressed as

$$L(\Theta) = \frac{1}{N}\sum_{i=1}^{N} \|F(P_i; \Theta) - illu_i\|^2. \tag{6}$$

where N is the number of image patch in a training batch, P_i represents a weakly illuminated image patch, $illu_i$ is a patch of illumination map that corresponds to P_i, and F is the learned mapping function.

3.3. Weakly illuminated image enhancement

Following previous method [11], we adjust the estimated illumination map by Gamma correction in order to thoroughly unveil dark regions in the results, which can be expressed as

$$L(x)' = L(x)^{\gamma}, \tag{7}$$

where $L(x)$ is the estimated illumination map and $L(x)'$ is the Gamma corrected illumination map. Here, $\gamma = 1.7$ is a heuristic value. We assume that the local region with size $n \times n$ of input image has the same illumination intensity when we optimize the mapping. Therefore, after Gamma correction, we need to refine the illumination map by guided image filtering [12] in order to remove the effects of blocking. In guided image filtering, the red channel of input image is taken as guided image and the size of filtering window is 16×16. Last, according to Eq. (1), the enhanced image $R(x)$ can be obtained by

$$R(x) = \frac{I(x)}{L_r(x)}. \tag{8}$$

where x denotes a pixel, $I(x)$ is the weakly illuminated image, and $L_r(x)$ is the refined illumination map. As shown in Fig. 2, our LightenNet accurately estimates the illumination maps which represent the illumination intensity. With the accurate illumination maps, our method generates natural and realistic results where dark regions are enhanced while light regions (*e.g.*, light source) are preserved.

3.4. LightenNet implementation

Our model is implemented in the Caffe package [14] and the LightenNet parameter settings are summarized in Table 1. In the stage of training LightenNet, the filter weights of each layer are initialized randomly from a Gaussian distribution, and the biases are set to 0. The initial learning rate is 0.05, and the learning rate decreases by 0.5 every 100,000 iterations. The momentum parameter is set to 0.9. A batch-mode learning method with a batch size of 128 is applied. The network training is done on a PC with a Intel(R) Xeon(R) CPU E5-2660 v3 @2.60 GHz and an Nvidia Titan X GPU within 1 h. Note that this optimization is real fast.

One challenge of training LightenNet is that deep learning systems require amount of training data, typically paired with labels or corresponding ground truth. Unfortunately, there is no enough labelled data available. Different from pervious methods which use the HDR image as labelled data, we synthesize training data based on the Retinex model in Eq. (1). Instead of taking HDR images as labels, our synthesis approach is physical formation model available, which results in more reasonable and reliable enhancement performance. In addition, to employ synthetic training data, we follow the assumptions that image content is independent of illumination map and illumination map in image patch is locally constant (*i.e.*, local region has the same illumination intensity).

Specifically, we collected 600 clear illuminated images with a variety of content from Internet for the synthesis of sample pairs (*i.e.*, pairs of weakly illuminated image and its illumination map). Here, clear illuminated images mean that images are with well illumination and contrast but without noise and blurring. please see the examples in Fig. 3. These 600 clear illuminated images have different sizes and are with PNG or BMP format. Based on Eq. (1) Retinex model, given a clear illuminated image $R(x)$ and a random illumination value L, a weakly illuminated image $I(x)$ can be synthesized as $I(x) = R(x) \cdot L$. Here, $R(x)$, L, and $I(x)$ are normalized. Based on these synthesized images, we obtain training image patches by overlapping cropping. The overlapping pixels are 10. At last, a total of 2,052,864 training image patches with size $n \times n = 16 \times 16$ are collected. Note that we have assumed that local region (*i.e.*, 16×16) has the same illumination intensity and the assumption is reasonable in the real-world when the image patch is small enough. Fig. 3 presents several examples of the clear illuminated images and the corresponding synthetic weakly illuminated images.

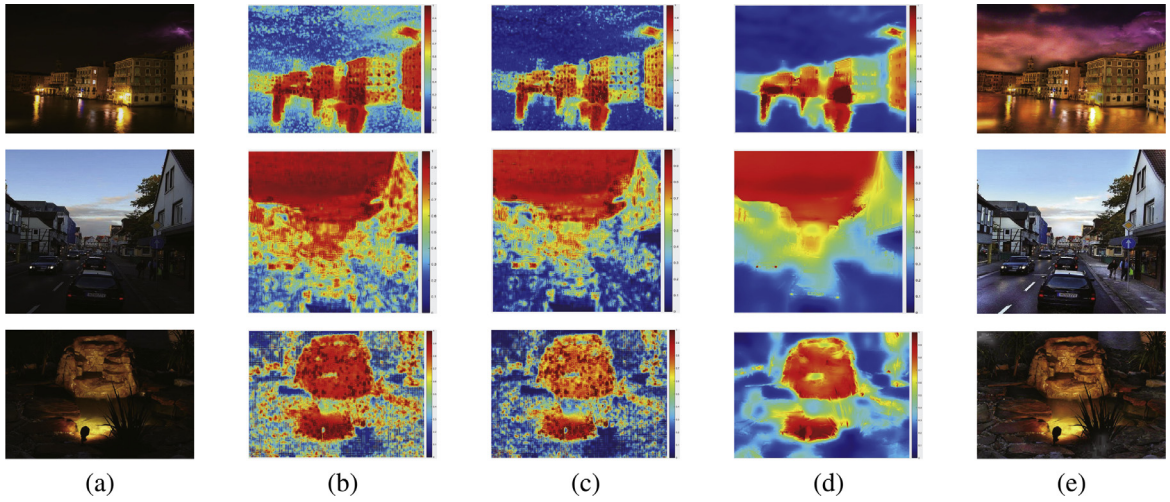


Fig. 2. Our step-by-step outputs. (a) Weakly illuminated images. (b) Coarse illumination maps estimated by LightenNet. (c) Gamma corrected illumination maps. (d) Refined illumination maps by guided image filtering. (e) Our enhanced results. In the (b)–(d), different color represents different illumination intensity (from blue to red represents from 0 to 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Parameter settings of LightenNet.

Formulation	Type	Input Size	Num	Filter	Padding
Patch Extraction and Representation	Conv	$3 \times 16 \times 16$	$n_1 = 32$	$f_1 \times f_1 = 6 \times 6$	0
	ReLU	$32 \times 11 \times 11$	32	–	0
Feature Enhancement	Conv	$32 \times 11 \times 11$	$n_2 = 16$	$f_2 \times f_2 = 7 \times 7$	0
	ReLU	$16 \times 5 \times 5$	16	–	0
Non-linear mapping	Conv	$16 \times 5 \times 5$	$n_3 = 8$	$f_3 \times f_3 = 1 \times 1$	0
	ReLU	$8 \times 5 \times 5$	8	–	0
Reconstruction	ConV	$8 \times 5 \times 5$	$n_4 = 1$	$f_4 \times f_4 = 5 \times 5$	0

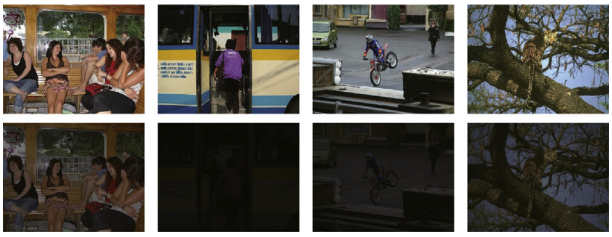


Fig. 3. Examples of the clear illuminated images and the corresponding synthetic weakly illuminated images. From top to bottom are the clear illuminated images and the corresponding synthetic weakly illuminated images. From left to right, the illumination values L are 0.5159, 0.0088, 0.2328, and 0.4475.

Table 2
Average MSE values for different network parameter settings and architectures.

Layers	Filters	Sizes	MSE
4	16-8-4-1	6-7-1-5	0.04225
4	**32-16-8-1**	**6-7-1-5**	**0.00816**
4	64-32-16-1	6-7-1-5	0.00623
4	128-64-32-1	6-7-1-5	0.00565
4	32-16-8-1	3-9-1-6	0.00865
4	32-16-8-1	5-7-1-6	0.04225
4	**32-16-8-1**	**6-7-1-5**	**0.00816**
4	32-16-8-1	7-5-1-6	0.00927
4	**32-16-8-1**	**6-7-1-5**	**0.00816**
4	32-16-8-1	6-7-3-5	0.00771
4	32-16-8-1	6-7-5-5	0.00750
4	32-16-8-1	6-7-7-5	0.00717
4	**32-16-8-1**	**6-7-1-5**	**0.00816**
8	32-16-8-1	6-7-1-5	0.00897
12	32-16-8-1	6-7-1-5	0.00923
16	32-16-8-1	6-7-1-5	0.01243

4. Experiment results

In this section, we first investigate the relations between the performance of LightenNet and network parameter settings. Next, we present the qualitative and quantitative comparisons on different types of weakly illuminated images. At last, we present several failure cases of our method and discuss the reasons.

4.1. Model and performance trade-offs

Based on the basic network parameter settings presented in Table 1, we will investigate the effects of different parameter settings and architectures. The MSE values for different settings are summarized in Table 2. The MSE value for our basic network parameter settings is marked in bold. The basic filter number settings of LightengNet are $n_1 = 32$, $n_2 = 16$, $n_3 = 8$, and $n_4 = 1$, and could be denoted as 32-16-8-1. Following this denotation, we denote another 3 filter number settings as 128-64-32-1, 64-32-16-1, and 16-8-4-1 to show the effects of filter number. Then, we present the effects of filter size. The basic filter size settings of our network are $f_1 = 6$, $f_2 = 7$, $f_3 = 1$, and $f_2 = 5$, and could be denoted as 6-7-1-5. Inspired by Sparse Representation-based method [5], we fix the filter size of third layer to be 1 (*i.e.*, $f_3 = 1$), and change filter size of other layers. We denote another 3 filter size settings as 7-5-1-6, 5-7-1-6, and 3-9-1-6. Besides, we also remove the 1×1 filter size constraint to investigate the effects of sparse-representation-based constraint. We denote another 3 filter size settings as 6-7-3-5, 6-7-5-5, and 6-7-7-5. At last, we investigate the effects of layer number. The default layer number is 4 (*i.e.*, patch extraction and representation layer, feature enhancement layer, non-linear mapping layer, and reconstruction layer). We repeat original architec-

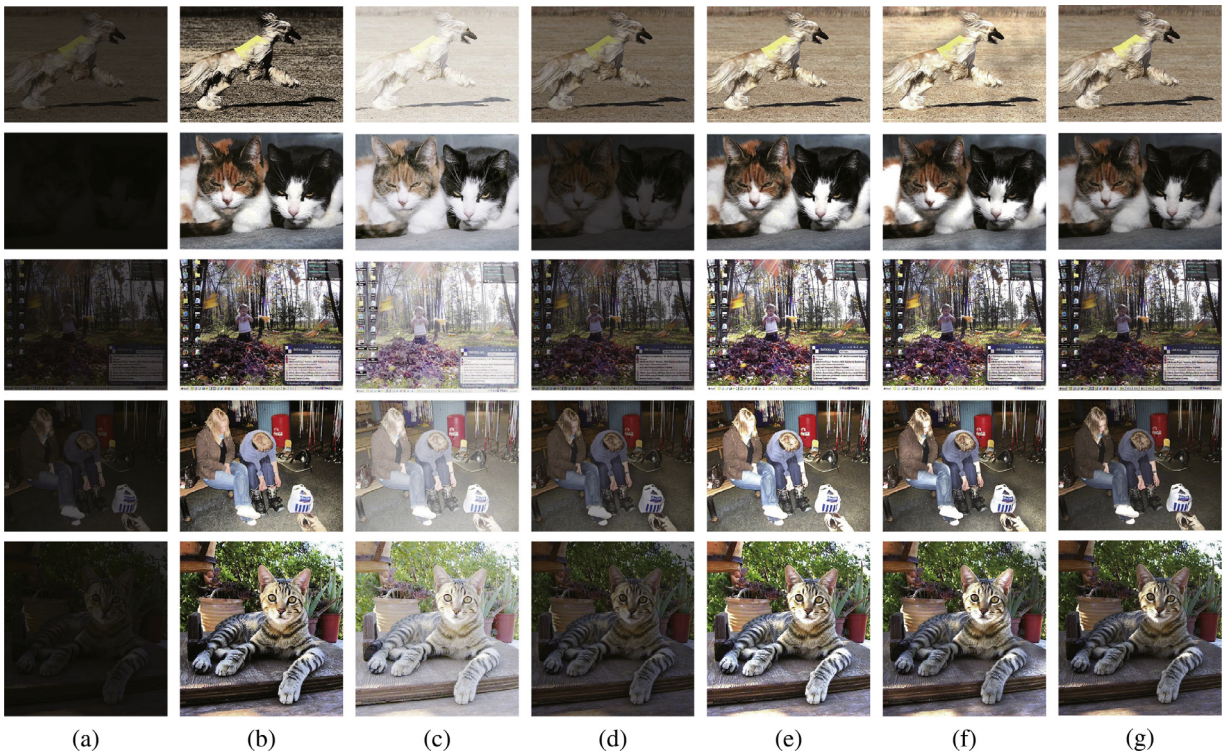


Fig. 4. Qualitative comparisons on synthetic images. (a) Synthetic weakly illuminated images. (b) Results of HE [10]. (c) Results of MSR [15]. (d) Results of AWVM [8]. (e) Results of LIME [11]. (f) Results of our method. (g) Ground truth.

ture several times in order to add more layers. 8, 12, and 16 represent repeating 2 times, 3 times, and 4 times original architecture.

As shown in Table 2, with filter number increasing, the superior MSE performance could be achieved. However, the improvement is limited. As we all known, the limited improvement is at the cost of training time and processing time. Thus, we select a middle network filter number as our basic filter number settings based on the trade off between performance and complexity. In addition, our basic filter size settings obtain best result among MSE comparisons. That may be because our filter size settings could grasp richer structural information, which in turn leads to better performance. It also demonstrates that a reasonable filter size settings could achieve better performance. For the 1×1 filter size constraint, we found that breaking this constraint could introduce the limited improvement at the cost of training time and processing time. Thus, we retain 1×1 constraint in our network since this constraint is effective and efficient for our network design. At last, with more layers, we do not obtain better MSE performance. The reason might be (1) the effects of gradient diffusion; (2) simple original architecture repeat results in unreasonable network architecture. In the future work, we will investigate more reasonable deep network architectures such as He et al. [13] and Kim et al. [16] for weakly illuminated image enhancement.

4.2. Qualitative comparisons

In this part, we qualitatively compare our method with classical HE (Histogram Equalization) method [10] and MSR (Multi-scale Retinex) method [15], and the state-of-the-art methods (*i.e.*, LIME [11] and AWVM [8]) on synthesized and real weakly illuminated images.

Firstly, we carry out qualitative comparisons on synthesized weakly illuminated images. These images are synthesized based on Retinex model in Eq. (1). Specifically, the synthesis approach has been illustrated in Section 3.4. Qualitative comparisons on synthetic images are presented in Fig. 4.

In Fig. 4, compared with synthetic weakly illuminated images, all of methods successfully improve the luminance and contrast. Compared to ground truth, the results of our method are most close to the ground truth of the weakly illuminated images. HE method increases the global contrast, which leads to over or under-enhanced regions. MSR method introduces gray veiling on the results because it completely removes the illumination. LIME method produces over-enhanced results while AWVM method tends to generate dim results. Next, we compare different methods on the real weakly illuminated images captured under varying illumination circumstances in Fig. 5.

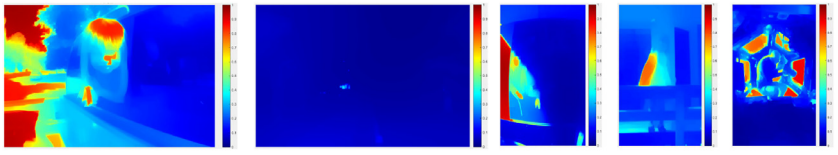
In Fig. 5, for the real weakly illuminated images, all of compared results have the same trend with those of synthetic images. Specifically, for images with non-uniform illumination inside images (*i.e.*, images "girl "group photo", and "girl2") where dark and bright areas coexist, HE and LIME methods produce over-enhanced and unrealistic results (*e.g.*, the hands of image "group photo" in the result of HE method and the clothing of image "girl2" in the result of LIME method). The proposed method produces natural appearance without over or under enhanced regions owe to the accurate estimation of illumination maps which are used to enhance the dark areas and maintain the bright areas. For the image with extremely low lighting (*i.e.*, image "tree"), HE and LIME methods unveil more details of scenario while they introduce artifacts and over-saturated regions (*e.g.*, the footpath and tree trunk). For the image with back lighting (*i.e.*, "man"), HE and LIME methods introduce artifacts (*e.g.*, the front regions) and over-enhanced regions. For both non-uniformly illuminated images and back lighting image, the results of MSR are gray and unrealistic because the illumination is completely removed. The results of AWVM look similar with our results which look pleasing and natural. Overall, the proposed method produces visual pleasing results for the weakly illuminated images taken under varying circumstances and the de-

(a) Real weakly illuminated images

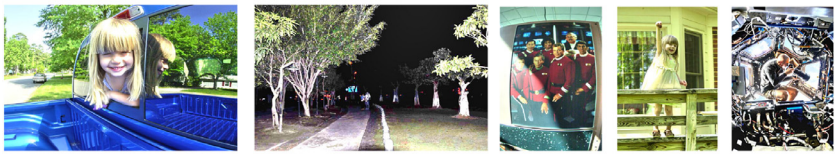
(b) Results of HE (Gonzalez and Woods, 2017)

(c) Results of MSR (Jobson et al., 1997)

(d) Results of LIME (Guo et al., 2017).

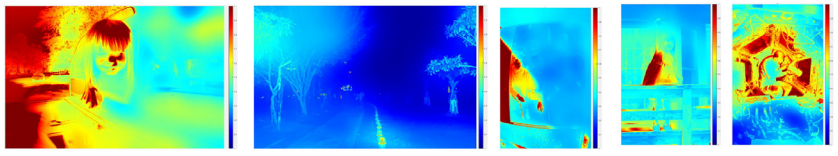


(e) Estimated illumination maps by LIME (Guo et al., 2017).



(f) Results of AWVM (Fu et al., 2016b).

(g) Results of our LightenNet.



(h) Estimated illumination maps by LightenNet.

Fig. 5. Qualitative comparisons on real images taken under varying illumination circumstances such as non-uniform illumination, extremely low lighting, and back lighting. From left to right are “girl1”, “trees”, “group photo”, “girl2”, and “man”.

Table 3
Quantitative results on synthetic images in terms of MSE, PSNR, and SSIM.

Method	MSE	PSNR	SSIM
HE	1.1623×10^3	19.5022	0.7829
MSR	6.7849×10^3	9.9812	0.6563
AWVM	2.9654×10^3	14.9027	0.7967
LIME	1.5017×10^3	18.2530	0.8620
Ours	$\mathbf{0.8827 \times 10^3}$	**21.7146**	**0.9257**

tails of scenes and objects are well restored. Observing the estimated illumination maps by LIME method and our method, it is obvious that LIME method tends to discard the details and textures of the scenes (*i.e.*, image "tree") when compared to our illumination maps, which leads to the suboptimum results of LIME method. It also illustrates that the illumination map estimated by CNN is more robust than conventional method.

4.3. Quantitative comparisons

To our best knowledge, there is no quantitative evaluation metric designed for weakly illuminated image enhancement methods. Hence, different researchers utilize different strategies to evaluate their results. Most of previous methods use the no-reference or full-reference image quality assessment metrics, such as Wang et al. [25], Xue et al. [27], and Peng et al. [22] to quantitatively assess their methods. However, it is unfair, especially for real weakly illuminated images, since these metrics do not take the characteristics of weakly illuminated image into account. Additionally, there is no real weakly illuminated database with ground truth available for full-reference image quality assessment metrics. Because most of low light image enhancement methods can achieve good results, it is hard to rank them visually. In order to fair compare different methods, we carry out two quantitative comparisons. For synthetic weakly illuminated images, we use MSE, Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity (SSIM) [26] to measure the differences between the results and ground truth. For real illuminated images, we do a user study on a dataset collected from Internet.

We collect 100 clear illuminated images, and then synthesize 100 weakly illuminated images based on Eq. (1) using the same approach with our network training data generation. Parts of synthetic images and the enhanced results have been shown in Fig. 4. In the comparisons on synthetic images, we compare the differences between the enhanced result and ground truth (*i.e.*, the corresponding clear illuminated image from Internet). Table 3 summarizes the average evaluation values in terms of MSE, PSNR, and SSIM for the results on the 100 synthetic images. In Table 3, the values in bold represent the best results.

As shown in Table 3, the proposed method stands out among the compared methods in terms of MSE, PSNR, and SSIM values. The best values indicate that the results of the proposed method are most close to the ground truth and demonstrate the effectiveness of the proposed method. In addition, it is interesting that the proposed method achieves the best performance on SSIM evaluation metric, although our LightenNet is optimized by the MSE loss function.

Furthermore, we conduct a user study on an image dataset which includes 40 real weakly illuminated images collected from Internet. These images are with a variety of content and light conditions (*e.g.*, well-lit, low light level, extremely low light level, non-uniform illumination, back lighting, and so on). Parts of collected images and the enhanced results have been shown in Fig. 5. The enhanced results are randomly displayed on a screen and separately scored by 7 participants who have image processing technical background according to visual quality. We repeat the scoring 5 times and summarize the average scores (from 0 (worst) to 100 (best)) for the results of different methods in Table 4.

Table 4
User study on real weakly illuminated images.

Method	HE	MSR	AWVM	LIME	Ours
Scores	74.10	32.59	50.21	82.88	**89.47**

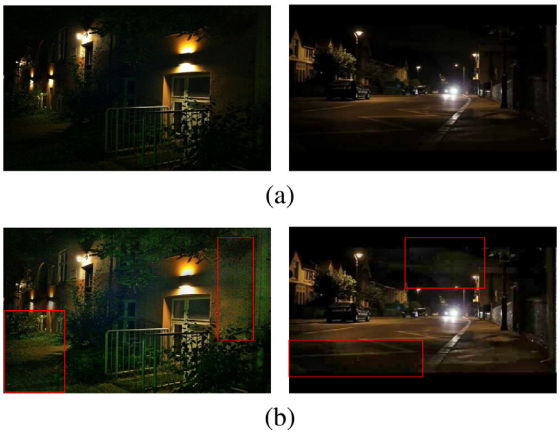


Fig. 6. Failure cases of our method. (a) Weakly illuminated images. From left to right are image captured by low quality device and image with JPG compression format. (b) Results of our method. The reader is encouraged to zoom in for a better view.

In Table 4, the user study results demonstrate that even for real weakly illuminated images captured under diverse illumination circumstances, the proposed method also can achieve visual pleasing results and is robust for different kinds of weakly illuminated images.

4.4. Failure cases

When training our LightenNet, we just consider the weakly illuminated images with good quality. Therefore, our method shows limitation when it is used to enhance weakly illuminated images with low quality such as noise and JPG compression. Fig. 6 presents several failure cases of our method.

In Fig. 6, our method amplifies the noise in the weakly illuminated image taken by the device are of low quality and introduces blocking artifacts for weakly illuminated image with JPG compression format. In the future work, we will take the effects of low quality weakly illuminated images into consideration when we train our network.

5. Conclusion

In this letter, we propose an effective CNN-based weakly illuminated image enhancement method. The proposed LightenNet learns a mapping between weakly illuminated image and the corresponding illumination map which is subsequently used to obtain the enhanced image. Our results are characteristics with natural and realistic appearance and improved brightness and contrast. The proposed method achieves superior performance than the state-of-the-art methods on both qualitative and quantitative comparisons.

Acknowledgments

This work was supported in part by the National Key Basic Research Program of China (2014CB340403), the National Natural Science Foundation of China (61771334), the program of China

Scholarship Council (201606250063), and the Australian Research Council grant (DP150104645).

References

[1] B. Cai, X. Xu, K. Jia, C. Qing, D. Tao, DehazeNet: an end-to-end system for single image haze removel, IEEE Trans. Image Process. 25 (2016) 5187–5198.
[2] C. Dong, Y. Deng, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach.Intell. 38 (2016) 295–307.
[3] C. Dong, Y. Deng, C.C. Loy, X. Tang, Compression artifacts reduction by a deep convolutional network, in: IEEE Conference on Computer Vision, 2015, pp. 576–584.
[4] X. Dong, G. Wang, Y. Pang, W. Li, J. Wen, Y. Lu, Fast efficient algorithm for enhancement of low lighting video, in: IEEE Conference on Multimedia and Expo, 2011, pp. 1–6.
[5] K. Fotiadou, G. Tsagkatakis, P. Tsakalides, Low light image enhancement via sparse representations, in: IEEE Conference on Image Analysis and Recognition, 2014, pp. 84–93.
[6] X. Fu, Y. Liao, D. Zeng, Y. Huang, X.P. Zhang, X. Ding, A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation, IEEE Trans. Image Process. 24 (2015) 4965–4977.
[7] X. Fu, D. Zeng, Y. Huang, X. Ding, J. Paisley, A fusion-based enhancing method for weakly illuminated images, Signal Process. 129 (2016) 82–96.
[8] X. Fu, D. Zeng, Y. Huang, X. Zhang, X. Ding, A weighted variational model for simultaneous reflectance and illumination estimation, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2782–2790.
[9] M. Gharbi, J. Chen, J.T. Barron, S.W. Hasinoff, F. Durand, Deep bilateral learning for real-time image enhancement, ACM Trans. Graph. 36 (2017).
[10] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Prentice–Hall, Englewood Cliffs, NJ, USA, 2017.
[11] X. Guo, Y. Li, H. Ling, Lime: low-light image enhancement via illumination map estimation, IEEE Trans. Image Process. 26 (2017) 982–993.
[12] K. He, J. Sun, X. Tang, Guided image filtering, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 1397–1409.
[13] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[14] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: convolutional architecture for fast feature embedding, in: IEEE Conference on Multimedia, 2014, pp. 675–678.
[15] D.J. Jobson, Z.U. Rahman, G.A. Woodell, A multi-scale Retinex for bridging the gap between color images and the human observation of scenes, IEEE Trans. Image Process. 6 (1997) 965–976.
[16] J. Kim, J.K. Lee, K.M. Lee, Accurate image super-resolution using very deep convolutional networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654.
[17] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. (2012).
[18] E. Land, An alternative technique for the computation of the designator in the Retinex theory of color vision, Natl. Acad. Sci. U 83 (1986) 3078–3080.
[19] L. Li, R. Wang, W. Wang, W. Gao, A low-light image enhancement method for both denosing and contrast enlarging, in: IEEE Conference on Image Processing, 2015, pp. 3730–3734.
[20] K.G. Lore, A. Akintayo, S. Sarkar, Llnet: a deep autoencoder approach to natural low-light image enhancement, Pattern Recognit. 61 (2017) 650–662.
[21] M.K. Ng, W. Wang, A total variation model for Retinex, SIAM J. Imaging Sci. 4 (2011) 345–365.
[22] Y. Peng, K. Jayant, K. Le, D. David, Unsuperived feature learning framework for non-reference image quality assessment, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1098–1105.
[23] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B.H. Romeny, J.B. Zimmerman, K. Zuiderveld, Adaptive histogram equalization and its variations, Comput. Vis. Graph. Image Process. 39 (1987) 355–368.
[24] L. Wang, L. Xiao, H. Liu, Z. Wei, Variational Bayesian method for Retinex, IEEE Trans. Image Process. 23 (2014) 3381–3396.
[25] S. Wang, J. Zheng, H.M. Liu, Naturalness preserved enhancement algorithm for non-uniform illumination images, IEEE Trans. Image Process. 22 (2013) 3538–3548.
[26] Z. Wang, A. Bovik, H. Sherikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612.
[27] W. Xue, L. Zhang, X. Mou, A. Bovik, Gradient magnitude similarity deviation: a highly efficient perceptual image quality index, IEEE Trans. Image Process. 23 (2014) 684–695.
[28] X. Zhang, P. Shen, L. Luo, L. Zhang, J. Song, Enhancement and noise reduction of very low light level images, in: IEEE Conference on Pattern Recognition, 2012, pp. 2034–2037.