

# Underexposed Photo Enhancement using Deep Illumination Estimation

Ruixing Wang<sup>1,\*</sup> Qing Zhang<sup>2,\*</sup> Chi-Wing Fu<sup>1</sup> Xiaoyong Shen<sup>3</sup> Wei-Shi Zheng<sup>2</sup> Jiaya Jia<sup>1,3</sup>  
<sup>1</sup>The Chinese University of Hong Kong <sup>2</sup>Sun Yat-sen University, China <sup>3</sup>YouTu Lab, Tencent

## Abstract

This paper presents a new neural network for enhancing underexposed photos. Instead of directly learning an image-to-image mapping as previous work, we introduce intermediate illumination in our network to associate the input with expected enhancement result, which augments the network’s capability to learn complex photographic adjustment from expert-retouched input/output image pairs. Based on this model, we formulate a loss function that adopts constraints and priors on the illumination, prepare a new dataset of 3,000 underexposed image pairs, and train the network to effectively learn a rich variety of adjustment for diverse lighting conditions. By these means, our network is able to recover clear details, distinct contrast, and natural color in the enhancement results. We perform extensive experiments on the benchmark MIT-Adobe FiveK dataset and our new dataset, and show that our network is effective to deal with previously challenging images.

## 1. Introduction

Photo sharing on social networks is very common due to the readily-available cameras on various devices, particularly the cell phones. However, the captured photo could get underexposed due to low- and back-lighting; see Figure 1(a) for an example. Such photos not only look unpleasing and fail to capture what the user desires, but also challenge many fundamental computer vision tasks, such as segmentation, object detection and tracking, since the underexposed regions have barely-visible details and relatively low contrast, as well as dull colors.

Severely-underexposed photo enhancement is a challenging task, since the underexposed regions are usually imperceptible and the enhancement process is highly non-linear and subjective. Although software exists to allow users to interactively adjust photos, it is rather tedious and difficult for non-experts since it requires to simultaneously manipulate controls like color and contrast, while finely tuning various objects and details in the photos. Several

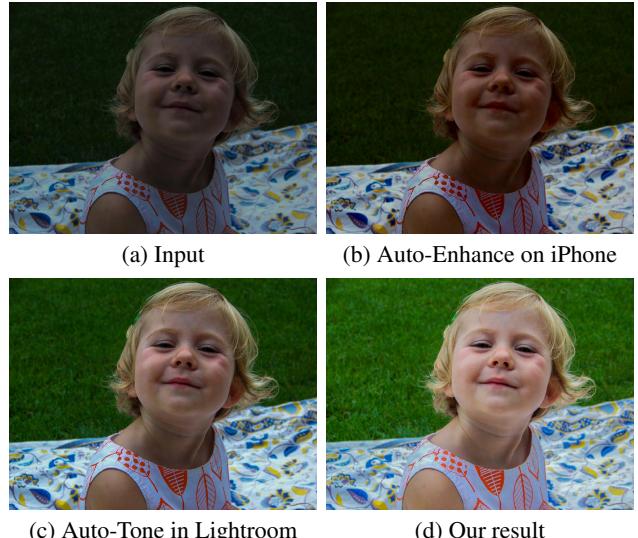


Figure 1: A challenging underexposed photo (a) enhanced by various tools (b)-(d). Our result contains more details, distinct contrast, and more natural color.

recent tools provide an automated function for users to enhance photos by just a single click, e.g., “Auto Enhance” on iPhone and “Auto Tone” in Lightroom. These tools do not greatly alter image contrast (and exposure) and may fail on severely underexposed images due to the inherent difficulty of automatically balancing assorted factors in the adjustment; see Figure 1.

On the other hand, various methods were proposed in the research community to tackle the problem. Early work [34, 25, 32, 11, 26, 4] primarily focuses on contrast enhancement, which may not be sufficient to recover image details and color. More recent work [16, 17, 13, 9, 15, 22] takes data-driven approaches to simultaneously learn adjustment in terms of color, contrast, brightness, and saturation for producing more expressive results. We note that existing methods still have their respective limitations on severely underexposed images; see Figure 2.

This paper presents a new end-to-end network for enhancing underexposed photos. Particularly, instead of directly learning an image-to-image mapping, we design our network to first estimate an *image-to-illumination mapping*

\*Joint first authors

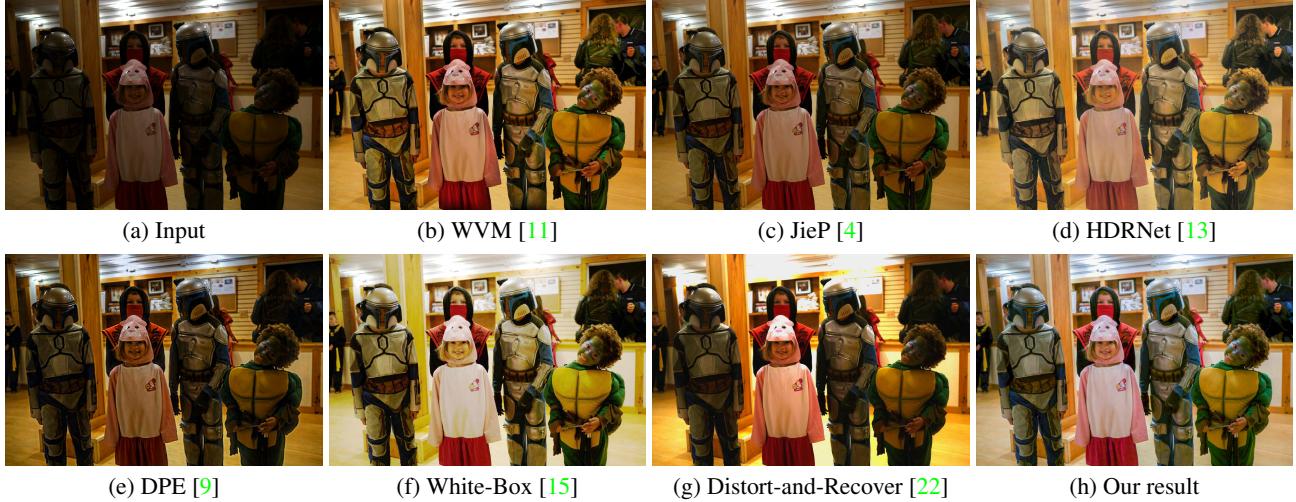


Figure 2: Another underexposed photo (a) enhanced by various methods (b)-(h). There exist unclear image details, distorted color, weak contrast, abnormal brightness, and unnatural white balance in various results.

for modeling varying-lighting conditions and then take the illumination map to light up the underexposed photo. By this approach, we make the learning process effective and *infer a rich variety of photographic adjustment*. Further, we adopt bilateral-grid-based upsampling to reduce the computational cost, and design a loss function that adopts various constraints and priors on illumination, so that we can efficiently recover underexposed photos with natural exposure, proper contrast, clear details, and vivid color. We also prepare a new dataset of 3,000 underexposed photos that cover diverse lighting conditions to supplement existing benchmark data. Below, we summarize the major contributions of this work.

- We propose a network for enhancing underexposed photos by estimating an image-to-illumination mapping, and design a new loss function based on various illumination constraints and priors.
- We prepare a new dataset of 3,000 underexposed images, each with an expert-retouched reference.
- We perform evaluation on our method using existing and new datasets, and demonstrate the superiority of our method qualitatively and quantitatively.

## 2. Related Work

Photo enhancement has a long history in computer vision and image processing. One pioneering method is the famous histogram equalization, which expands the dynamic range and increases image contrast. Its limitation is obvious with the globally adjusted contrast in the entire image.

**Retinex-based Methods** Assuming that an image can be decomposed into pixel-wise product of reflectance and illumination (or shading), Retinex-based methods [19] treat

the reflectance component as a plausible approximation to the enhanced image. Hence, photo enhancement can be formulated as an illumination estimation problem, where illumination is estimated to enhance the underexposed photos [27, 11, 31, 14, 4, 33]. However, due to the nonlinearity across color channels and data complexity, existing methods have limited capability to enhance color, since color is easily distorted locally. Our work also considers illumination estimation and yet it advances state of the arts in two aspects. First, the neural network learns the illumination by exploiting massive photos in diverse lighting conditions and models a rich variety of photographic adjustment. Second, our approach enables nonlinear color enhancement from multi-channel illumination.

**Learning-based Methods** Recent effort on photo enhancement is mostly learning-based. For instance, Bychkovsky *et al.* [3] provided the first and largest dataset MIT-Adobe FiveK with input and expert-retouched image pairs for tone adjustment. Yan *et al.* [28] presented a machine-learned ranking approach for automatically enhancing color in a photograph. Yan *et al.* [29] constructed the semantic map to achieve semantic-aware photo enhancement. Lore *et al.* [21] proposed a deep autoencoder-based approach for low-light image enhancement, while Gharbi *et al.* [13] introduced bilateral learning for real-time performance. Yang *et al.* [30] corrected the LDR images using a deep reciprocating HDR transformation. Cai *et al.* [5] learned a contrast enhancer from multi-exposure images. Recently, Chen *et al.* [9] developed an unpaired learning model for photo enhancement based on a two-way generative adversarial networks (GANs), while Ignatov *et al.* [18] designed a weakly-supervised image-to-image GAN-based network. Further, Deng *et al.* [10] enabled aesthetic-driven

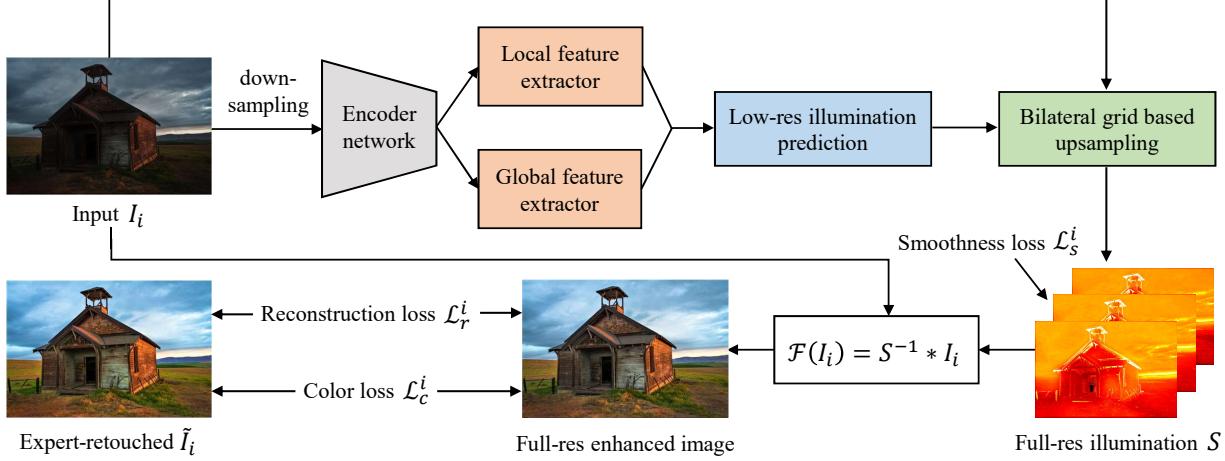


Figure 3: Overview of our network. First, we downsample and encode the input into a feature map, extract local and global features, and concatenate them to predict the low-res illumination via a convolution layer. Then we upsample the result to produce the full-res multi-channel illumination  $S$  (hot color map), and take it to recover the full-res enhanced image. We train the end-to-end network to learn  $S$  from image pairs  $\{I_i, \tilde{I}_i\}$  with three loss components  $\{L_r^i, L_s^i, L_c^i\}$ .

image enhancement by adversarial learning, while Chen *et al.* [6] addressed extreme low-light imaging by operating directly on raw sensor data with a new dataset.

Reinforcement learning was also employed to enhance the image adjustment process [15, 22]. Our approach is complementary to existing learning-based methods in two ways. First, we estimate the illumination mapping, unlike others that are based on image-to-image regression. Second, our new dataset exactly suits underexposed photo enhancement, which supplements other benchmark datasets and provides more real-world examples in diverse lighting conditions.

### 3. Methodology

#### 3.1. Image Enhancement Model

Fundamentally, the image enhancement task can be regarded as seeking a mapping function  $\mathcal{F}$ , such that  $\tilde{I} = \mathcal{F}(I)$  is the desired image, enhanced from input image  $I$ . In recent Retinex-based image enhancement methods [11, 14], the inverse of  $\mathcal{F}$  is typically modeled as an illumination map  $S$ , which multiplies with the reflectance image  $\tilde{I}$  in a pixel-wise manner to produce the observed image  $I$ :

$$I = S * \tilde{I}, \quad (1)$$

where  $*$  denotes a pixel-wise multiplication.

Similar to that of [11, 14], we also regard the reflectance component  $\tilde{I}$  as a well-exposed image, so in our model, we take  $\tilde{I}$  as the enhancement result and  $I$  as the observed underexposed image. Once  $S$  is known, we can obtain the enhancement result  $\tilde{I}$  by  $\mathcal{F}(I) = S^{-1} * I$ . Unlike existing work [11, 14], we model  $S$  as a multi-channel (R, G,

B) data instead of single-channel one to increase its competence in modeling color enhancement, especially for handling the nonlinearity across different color channels.

**Why this Model Works?** By introducing intermediate illumination in our network, we train the network to learn an image-to-illumination (instead of image-to-image) mapping. The key advantage is that illumination maps for natural images typically have relatively simple forms with known priors. So the network can have stronger generalization capability and be trained effectively to learn complex photographic adjustment for diverse lighting conditions. In addition, the model enables customizing the enhancement results by formulating constraints on illumination. For instance, contrast can be enhanced by enforcing locally smooth illumination, or setting the preferred exposure level by constraining illumination magnitudes.

#### 3.2. Network Architecture

Figure 3 presents the pipeline of our network, with the two major advantages of effective learning of the illumination mapping and efficient network computation.

**Effective Learning** Enhancing underexposed photos requires adjusting both local (*e.g.*, contrast, detail sharpness, shadow, and highlight) and global features (*e.g.*, color distribution, average brightness, and scene category). We consider local and global context from the features generated from an encoder network; see Figure 3 (top). To drive the network to learn the illumination mapping from the input underexposed image ( $I_i$ ) and corresponding expert-retouched image ( $\tilde{I}_i$ ), we design a loss function, with a smoothness prior on the illumination and a reconstruction

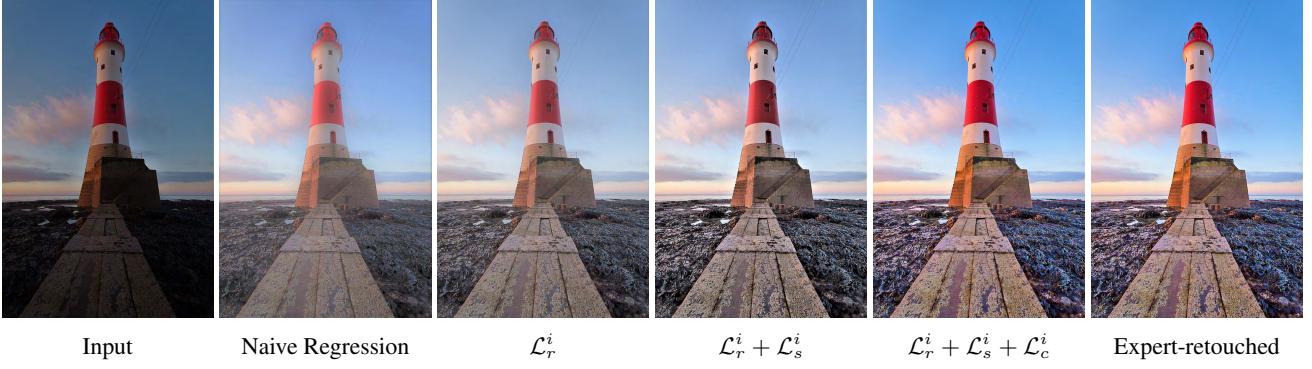


Figure 4: Ablation study that demonstrates the effectiveness of each component ( $\mathcal{L}_r^i$ ,  $\mathcal{L}_s^i$ , and  $\mathcal{L}_c^i$ ) in the loss function.

and color loss on the enhanced image; see Figure 3 (bottom). These strategies effectively learn  $S$  from  $(I_i, \tilde{I}_i)$  for recovering the enhanced image with a rich variety of photographic adjustment.

**Efficient Runtime** We learn the local and global features for predicting the image-to-illumination mapping in low resolution, and perform bilateral grid based upsampling [8, 7, 12, 13] to enlarge the low-res prediction to the full resolution; see Figure 3. Hence, most network computation is done in low-res domain, enabling real-time processing of high-resolution images.

### 3.3. Loss Function

We learn the illumination mapping from a set of  $N$  image pairs  $\{(I_i, \tilde{I}_i)\}_{i=1}^N$ . It produces  $S$  and the enhancement result  $\mathcal{F}(I) = S^{-1} * I$ . We design a loss function  $\mathcal{L}$  that consists of three components and minimize it during the network training. It is expressed as

$$\mathcal{L} = \sum_{i=1}^N \omega_r \mathcal{L}_r^i + \omega_s \mathcal{L}_s^i + \omega_c \mathcal{L}_c^i, \quad (2)$$

where  $\mathcal{L}_r^i$ ,  $\mathcal{L}_s^i$  and  $\mathcal{L}_c^i$  are the loss components, and  $\omega_r$ ,  $\omega_s$  and  $\omega_c$  are the corresponding weights. Note that we empirically set  $\omega_r=1$ ,  $\omega_s=2$ , and  $\omega_c=1$ .

**Reconstruction Loss** To obtain the predicted illumination  $S$ , we define the  $L_2$  error metric to measure the reconstruction error as

$$\begin{aligned} \mathcal{L}_r^i &= \|I_i - S * \tilde{I}_i\|^2, \\ \text{s.t. } (I_i)_c &\leq (S)_c \leq 1, \quad \forall \text{ pixel channel } c, \end{aligned} \quad (3)$$

where all pixel channels in  $I_i$  and  $\tilde{I}_i$  are normalized to  $[0,1]$ ,  $(\cdot)_c$  denotes a pixel color channel, and  $(I_i)_c \leq (S)_c \leq 1$  is the multi-channel illumination range constraint. Since  $\mathcal{F}(I_i) = S^{-1} * I_i$ , setting  $I_i$  as  $S$ 's lower bound ensures all color channels in the enhancement result  $\mathcal{F}(I_i)$  are (upper) bounded by one, thus avoiding colors

beyond the gamut, whereas setting 1 as  $S$ 's upper bound avoids mistakenly darkening the underexposed regions.

Figure 4 presents ablation study results that demonstrate the effect of various components in the loss function. Comparing the 2nd and 3rd images in the figure, we observe clearer details and better contrast in the result by minimizing the reconstruction loss. It has clear advantages over naive image-to-image regression, where the latter directly regresses the output image without estimating the intermediate illumination in our network (see Figure 3). While images enhanced with the reconstruction loss look more similar to the expert-retouched ones, there is still risk of not producing correct contrast details and vivid color (the 3rd and 6th images in Figure 4). Hence, we also introduce the smoothness and color loss.

**Smoothness Loss** According to the smoothness prior [23, 20, 2], illumination in natural images is in general locally smooth. Adopting this prior in our network has two advantages. First, it helps reduce over-fitting and increase the network's generalization capability. Second, it enhances the image contrast. When adjacent pixels  $p$  and  $q$  have similar illumination values, their contrast in the enhanced image can be estimated as  $|\tilde{I}_p - \tilde{I}_q| \approx S_p^{-1} * |I_p - I_q|$ , which should also be enlarged, since  $S \leq 1$ . Therefore, we define the smoothness loss on the predicted full resolution illumination  $S$  in Figure 3 as

$$\mathcal{L}_s^i = \sum_p \sum_c \omega_{x,c}^p (\partial_x S_p)_c^2 + \omega_{y,c}^p (\partial_y S_p)_c^2, \quad (4)$$

where we sum over all channels ( $c$ ) of all pixels;  $\partial_x$  and  $\partial_y$  are partial derivatives in horizontal and vertical directions in the image space; and  $\omega_{x,c}^p$  and  $\omega_{y,c}^p$  are spatially-varying (per-channel) smoothness weights expressed as

$$\begin{aligned} \omega_{x,c}^p &= (|\partial_x L_i^p|_c^\theta + \epsilon)^{-1} \text{ and} \\ \omega_{y,c}^p &= (|\partial_y L_i^p|_c^\theta + \epsilon)^{-1}. \end{aligned} \quad (5)$$

Here,  $L_i$  is the logarithmic image of the input image  $I_i$ ;  $\theta = 1.2$  is a parameter that controls the sensitivity to image



Figure 5: Example images in our dataset. **Top:** input. **Bottom:** corresponding expert-retouched reference images.

gradients; and  $\epsilon$  is a small constant typically set to 0.0001, preventing division by zero.

Intuitively, the smoothness loss encourages the illumination to be smooth on pixels with small gradients and discontinuous on pixels with large gradients. It is intriguing to note that for underexposed photos, image content and details are often weak. Large gradients are more likely incurred by inconsistent illumination. As demonstrated by the 4th image in Figure 4, by further incorporating the smoothness loss, we recover decent image contrast and clearer details compared with the results produced with only the reconstruction loss.

**Color Loss** Next, we formulate the color loss to encourage the color in the generated image  $\mathcal{F}(I_i)$  of  $I_i$  to match that in the corresponding expert-retouched image  $\tilde{I}_i$  as

$$\mathcal{L}_c^i = \sum_p \angle((\mathcal{F}(I_i))_p, (\tilde{I}_i)_p), \quad (6)$$

where  $(\cdot)_p$  denotes a pixel;  $\angle(\cdot, \cdot)$  is an operator that calculates the angle between two colors regarding the RGB color as a 3D vector. Eq. (6) sums the angles between the color vectors for every pixel pair in  $\mathcal{F}(I_i)$  and  $\tilde{I}_i$ .

The reasons that we use this simple formulation instead of an  $L_2$  distance in other color space are as follows. First, the reconstruction loss has already implicitly measured the  $L_2$  color difference. Second, since the  $L_2$  metric only numerically measures the color difference, it cannot ensure that the color vectors have the same direction. Therefore, the metric may induce evident color mismatch. This can be observed by comparing the 4th and 5th results with and without the color loss in Figure 4. Last but not least, the formulation is simple and fast for network computation.

### 3.4. Training Dataset

We prepared a new dataset of 3,000 images. We trained our network on it instead of the MIT-Adobe FiveK dataset [3] for two reasons. First, the FiveK dataset was created primarily for enhancing general photos rather than underexposed photos; it contains only a very small portion

(around 4%) of underexposed images. Second, the underexposed images in the benchmark dataset cover limited lighting conditions; it lacks challenging cases such as nighttime images and images with non-uniform illumination.

To prepare our dataset, we first capture images in the resolution of  $6000 \times 4000$  using Canon EOS 5D Mark III and Sony ILCE-7, and further collected around 15% more images from Flickr by searching with keywords “underexposed”, “low-light”, and “backlit”. Then, we recruited three experts from the school of photography to prepare a retouched reference image for each collected image using Adobe Lightroom. Our dataset is diverse; it covers a broad range of lighting conditions, scenes, subjects, and styles. Please see Figure 5 for some of the image pairs. Finally, we randomly split the images in the dataset into two subsets: 2,750 images for training and the rest for testing.

### 3.5. Implementation Details

We build our network on TensorFlow [1] and train it for 40 epochs with a mini-batch size of 16 on an NVidia Titan X Pascal GPU. The entire network is optimized using the Adam optimizer with a fixed learning rate of  $10^{-4}$ . For data augmentation, we randomly cropped  $512 \times 512$  patches followed by random mirror, resize and rotation for all patches. The downsampled input has a fixed resolution of  $256 \times 256$ . The encoder network is a pre-trained VGG16 [24]. The local feature extractor contains two convolution layers, while the global feature extractor contains two convolution layers and three fully-connected layers. Further, we use the bilateral grid-based module [13] to upsample the output. Our code and dataset are available at <https://github.com/wangruixing/DeepUPE>.

## 4. Experimental Results

**Datasets** We evaluated our network on (i) our dataset and (ii) the MIT-Adobe FiveK [3] dataset with 5,000 raw images, each with five retouched images produced by different experts (A/B/C/D/E). For the MIT-Adobe FiveK dataset, we follow previous methods [13, 15, 22] to use only the output

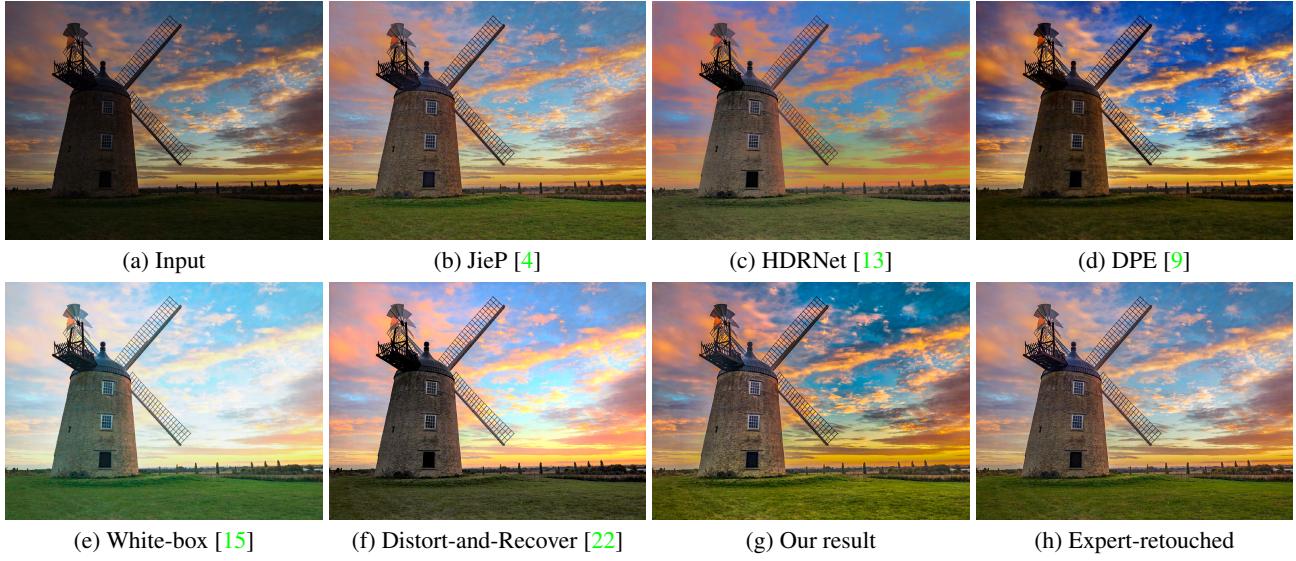


Figure 6: Visual comparison with state-of-the-art methods on a test image (a) from our dataset.

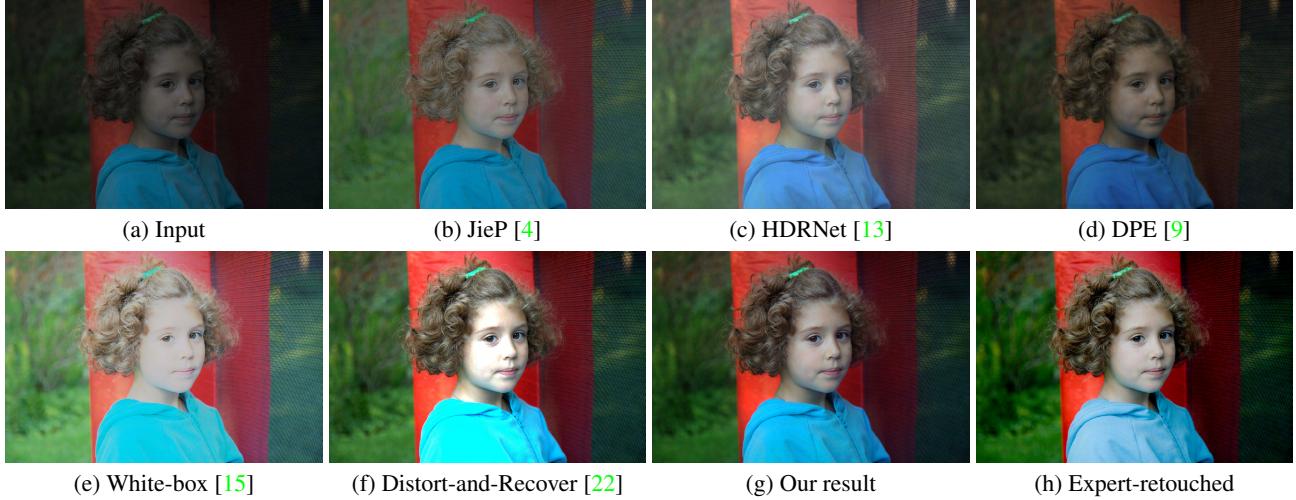


Figure 7: Visual comparison with state-of-the-art methods on a test image (a) from the MIT-Adobe FiveK [3] dataset.

by Expert C, randomly selected 500 images for validation and testing, and trained on the remaining 4,500 images.

**Evaluation Metrics** We employed two commonly-used metrics (i.e., PSNR and SSIM) to quantitatively evaluate the performance of our network in terms of the color and structure similarity between the predicted results and the corresponding expert-retouched images. Although it is not absolutely indicative, in general, high PSNR and SSIM values correspond to reasonably good results.

#### 4.1. Comparison with State-of-the-art Methods

We compare our method with the following five state-of-the-art image enhancement methods: (i) the latest Retinex-based method, JieP [4], and (ii)-(v) four recent deep-learning-based methods of HDRNet [13], DPE [9], White-

Box [15], and Distort-and-Recover [22]. For fair comparison, we produce their results using publicly-available implementation provided by the authors with recommended parameter setting. For the four learning-based methods, we further re-train their models on our dataset and also on the MIT-Adobe FiveK dataset to produce the best possible results. Our comparison is threefold.

**Visual Comparison** First, we show visual comparison in Figures 6 and 7 on two challenging cases using an unevenly-exposed photo with imperceptible windmill details (from our dataset) and an overall low-light photo with little portrait details (from the MIT-Adobe FiveK dataset). Comparing the results, we notice two key improvements of our method (h) over the others (b)-(f). First, our method is able to recover more details and better contrast in both

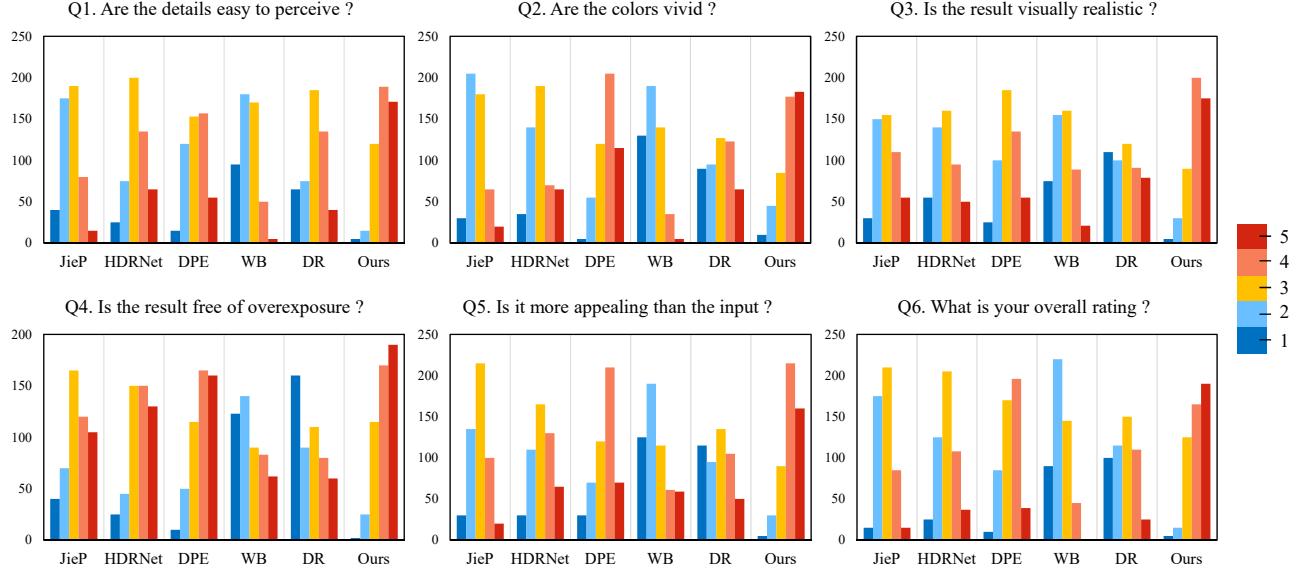


Figure 8: Rating distributions for various methods on the six questions in the user study. The ordinate axis shows the rating frequency received by the methods from the participants. WB and DR mean White-Box [15] and Distort-and-Recover [22].

Table 1: Quantitative comparison between our method and state-of-the-art methods on our dataset (w/o - without).

Method	PSNR	SSIM
HDRNet [13]	26.33	0.743
DPE [9]	23.58	0.737
White-Box [15]	21.69	0.718
Distort-and-Recover [22]	24.54	0.712
Ours w/o $\mathcal{L}_r$ , w/o $\mathcal{L}_s$ , w/o $\mathcal{L}_c$	27.02	0.762
Ours with $\mathcal{L}_r$ , w/o $\mathcal{L}_s$ , w/o $\mathcal{L}_c$	28.97	0.783
Ours with $\mathcal{L}_r$ , with $\mathcal{L}_s$ , w/o $\mathcal{L}_c$	30.03	0.822
Ours	<b>30.97</b>	<b>0.856</b>

Table 2: Quantitative comparison between our method and state-of-the-arts on the MIT-Adobe FiveK dataset.

Method	PSNR	SSIM
HDRNet [13]	28.61	0.866
DPE [9]	24.66	0.850
White-Box [15]	23.69	0.701
Distort-and-Recover [22]	28.41	0.841
Ours w/o $\mathcal{L}_r$ , w/o $\mathcal{L}_s$ , w/o $\mathcal{L}_c$	28.81	0.867
Ours with $\mathcal{L}_r$ , w/o $\mathcal{L}_s$ , w/o $\mathcal{L}_c$	29.41	0.871
Ours with $\mathcal{L}_r$ , with $\mathcal{L}_s$ , w/o $\mathcal{L}_c$	30.71	0.884
Ours	<b>30.80</b>	<b>0.893</b>

foreground and background, without obviously sacrificing over/underexposing part of the image. Second, it also reveals vivid and natural color, making the enhanced results look more realistic. Please see the supplementary material for more visual comparison results.

**Quantitative Comparison** To evaluate the learning effectiveness and generalization capability of our network, we quantitatively compare it with the other methods using the PSNR and SSIM metrics. Tables 1 and 2 report the results, where for each case, we re-trained our network, as well as others', on respective datasets. Note that our loss function without  $\mathcal{L}_r$ ,  $\mathcal{L}_s$ , and  $\mathcal{L}_c$  reduces to a pixel-wise  $L_2$  loss between corresponding image pairs in the dataset. Here, we do not include JieP [4] because it is not a learning-based method. For both comparisons, our method performs better, manifesting that our method not only effectively learns the photographic adjustment for enhancing the underexposed photos but also well generalizes to the MIT-Adobe FiveK dataset with a limited amount of underexposed photos.

**User Study** Further, we conducted a user study with 500 participants to compare results. Akin to that of [22], we first crawl 100 test images, which have over 50% pixels with intensity lower than 0.3, from Flickr by searching with keywords “city”, “flower”, “food”, “landscape”, and “portrait” (Figure 9 has an example). Then, we enhance each test image using our method and others', and recruit participants via Amazon Mechanical Turk to rate each group of results, which are presented in a random order to avoid subjective bias.

For each result, the participants are asked to give a rating for each of the six questions shown in Figure 8 using a Likert scale from 1 (worst) to 5 (best). Figure 8 summarizes the results, where each subfigure shows six rating distributions of the methods on a particular question. The distribution across methods shows that our results are more preferred by human subjects, where our method receives more “red” and far less “blue” ratings compared to the others. We also



Figure 9: Visual comparison with state-of-the-art methods on a test image employed in our user study.

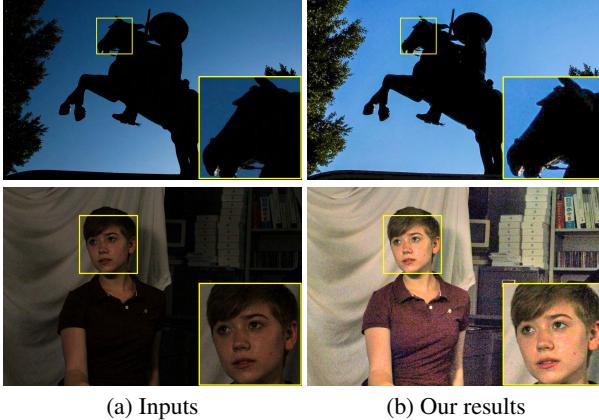


Figure 10: Failure cases. Input images with mostly black regions (*top row*) and with noise (*bottom row*).

performed a statistical analysis on the ratings by conducting paired t-test between our method and others. The result is clear: all the t-test results are statistically significant with  $p < 0.01$ . Please see the supplementary material for more details. Moreover, we extend the user study to compare also with “Auto Enhance” on iPhone and “Auto Tone” in Lightroom. Results are also contained in the supplementary material.

## 4.2. Discussions

**Ablation Study** Besides the visual results shown in Figure 4, we quantitatively evaluate the effectiveness of the components in our method. Comparing the statistics in the last row (ours) and the 5th row (ours without all three losses) in Tables 1 and 2, we observe clear advantage of our method in learning an image-to-illumination mapping over a naive image-to-image mapping. Moreover, the last four rows in each table reveal progressive improvement on the results by

having more loss components in our method, for both the MIT-Adobe FiveK dataset and our dataset. They convincingly demonstrate the effectiveness of each loss component.

**Limitations** Figure 10 presents two examples where our method, as well as other state-of-the-arts, all fail to produce visually compelling results. For the top image, we fail to recover details on the horse body, since the region is almost black without any trace of texture in the original image, while for the bottom input, our method does not clear noise in the enhancement result. Thus stronger denoising ability will be our future goal.

## 5. Conclusion

We have presented a new end-to-end network for enhancing underexposed photos. Our key idea is to learn an image-to-illumination (instead of image-to-image) mapping, so as to leverage the simplicity nature of illumination in natural images for the network to effectively learn a rich variety of photographic adjustment. Further, we design a loss function that adopts various constraints and priors on illumination and create a new dataset of 3,000 underexposed image pairs, enabling our network to recover clear details, distinct contrast, and vivid colors in underexposed photos. We have performed extensive experiments on our dataset and the MIT-Adobe FiveK dataset, and compared our method with five state-of-the-art methods to show the superiority of our solution in terms of visual comparison, quantitative comparison in terms of the PSNR and SSIM metrics, and a user study with 500 participants involved.

Our future work is to incorporate a denoising module into our network and extend our method to handling videos. Another direction is to address the nearly black regions by leveraging techniques in scene semantic analysis and photographic image synthesis.

## References

- [1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. TensorFlow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016. 5
- [2] Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8):1670–1687, 2015. 4
- [3] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédéric Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR*, 2011. 2, 5, 6
- [4] Bolun Cai, Xianming Xu, Kailing Guo, Kui Jia, Bin Hu, and Dacheng Tao. A joint intrinsic-extrinsic prior model for retinex. In *ICCV*, 2017. 1, 2, 6, 7, 8
- [5] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. 2
- [6] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, 2018. 3
- [7] Jiawen Chen, Andrew Adams, Neal Wadhwa, and Samuel W Hasinoff. Bilateral guided upsampling. *ACM Trans. on Graphics (SIGGRAPH Asia)*, 35(6):203, 2016. 4
- [8] Jiawen Chen, Sylvain Paris, and Frédéric Durand. Real-time edge-aware image processing with the bilateral grid. 26(3):103, 2007. 4
- [9] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs. In *CVPR*, 2018. 1, 2, 6, 7, 8
- [10] Yubin Deng, Chen Change Loy, and Xiaoou Tang. Aesthetic-driven image enhancement by adversarial learning. In *MM. ACM*, 2018. 2
- [11] 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. 1, 2, 3, 8
- [12] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédéric Durand. Deep joint demosaicking and denoising. *ACM Trans. on Graphics (SIGGRAPH Asia)*, 35(6):191, 2016. 4
- [13] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédéric Durand. Deep bilateral learning for real-time image enhancement. *ACM Trans. on Graphics (SIGGRAPH)*, 36(4):118, 2017. 1, 2, 4, 5, 6, 7, 8
- [14] 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, 3
- [15] Yuanming Hu, Hao He, Chenxi Xu, Baoyuan Wang, and Stephen Lin. Exposure: A white-box photo post-processing framework. *ACM Trans. on Graphics (SIGGRAPH)*, 37(2):26, 2018. 1, 2, 3, 5, 6, 7, 8
- [16] Sung Ju Hwang, Ashish Kapoor, and Sing Bing Kang. Context-based automatic local image enhancement. In *ECCV*, 2012. 1
- [17] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *ICCV*, 2017. 1
- [18] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. WESPE: Weakly supervised photo enhancer for digital cameras. In *CVPR*, 2018. 2
- [19] Edwin H Land. The retinex theory of color vision. *Scientific American*, 237(6):108–129, 1977. 2
- [20] Yu Li and Michael S Brown. Single image layer separation using relative smoothness. In *CVPR*, 2014. 4
- [21] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. LInet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 2
- [22] Jongchan Park, Joon-Young Lee, Donggeun Yoo, and In So Kweon. Distort-and-recover: Color enhancement using deep reinforcement learning. In *CVPR*, 2018. 1, 2, 3, 5, 6, 7, 8
- [23] Carsten Rother, Martin Kiefel, Lumin Zhang, Bernhard Schölkopf, and Peter V Gehler. Recovering intrinsic images with a global sparsity prior on reflectance. In *NIPS*, 2011. 4
- [24] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 5
- [25] J Alex Stark. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Transactions on Image Processing*, 9(5):889–896, 2000. 1
- [26] Shuhang Wang, Woon Cho, Jinbeum Jang, Mongi A Abidi, and Joonki Paik. Contrast-dependent saturation adjustment for outdoor image enhancement. *JOSA A*, 34(1):7–17, 2017. 1
- [27] 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
- [28] Jianzhou Yan, Stephen Lin, Sing Bing Kang, and Xiaoou Tang. A learning-to-rank approach for image color enhancement. In *CVPR*, 2014. 2
- [29] Zhicheng Yan, Hao Zhang, Baoyuan Wang, Sylvain Paris, and Yizhou Yu. Automatic photo adjustment using deep neural networks. *ACM Trans. on Graphics*, 35(2):11, 2016. 2
- [30] Xin Yang, Ke Xu, Yibing Song, Qiang Zhang, Xiaopeng Wei, and Rynson WH Lau. Image correction via deep reciprocating hdr transformation. In *CVPR*, 2018. 2
- [31] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new low-light image enhancement algorithm using camera response model. In *ICCV*, 2017. 2
- [32] Lu Yuan and Jian Sun. Automatic exposure correction of consumer photographs. In *ECCV*, 2012. 1
- [33] Qing Zhang, Ganzhao Yuan, Chunxia Xiao, Lei Zhu, and Wei-Shi Zheng. High-quality exposure correction of under-exposed photos. In *MM. ACM*, 2018. 2
- [34] Karel Zuiderveld. Contrast limited adaptive histogram equalization. *Graphics Gems*, pages 474–485, 1994. 1