Self-Supervised Image Enhancement and Denoising

Authors: Zhang Yu

Date: 2021-03-01T00:00:00+00:00

Abstract

This paper proposes a self-supervised low light image enhancement method based on deep learning, which can improve the image contrast and reduce noise at the same time to avoid the blur caused by pre-/post-denoising. The method contains two deep sub-networks, an Image Contrast Enhancement Network (ICE-Net) and a Re-Enhancement and Denoising Network (RED-Net). The ICE-Net takes the low light image as input and produces a contrast enhanced image. The RED-Net takes the result of ICE-Net and the low light image as input, and can re-enhance the low light image and denoise at the same time. Both of the networks can be trained with low light images only, which is achieved by a Maximum Entropy based Retinex (ME-Retinex) model and an assumption that noises are independently distributed. In the ME-Retinex model, a new constraint on the reflectance image is introduced that the maximum channel of the reflectance image conforms to the maximum channel of the low light image and its entropy should be the largest, which converts the decomposition of reflectance and illumination in Retinex model to a non-ill-conditioned problem and allows the ICE-Net to be trained with a self-supervised way. The loss functions of RED-Net are carefully formulated to separate the noises and details during training, and they are based on the idea that, if noises are independently distributed, after the processing of smoothing filters (e.g. mean filter), the gradient of the noise part should be smaller than the gradient of the detail part. It can be proved qualitatively and quantitatively through experiments that the proposed method is efficient.

Full Text

Preamble

Self-supervised Low Light Image Enhancement and Denoising

Yu Zhang, Xiaoguang Di, Bin Zhang, Chunhui Wang
Department of Electronic Science and Technology, Harbin Institute of Technology
zhangyuhit2@hit.edu.cn

Qingyan Li, Shiyu Yan

Abstract

This paper proposes a self-supervised low-light image enhancement method based on deep learning that simultaneously improves image contrast and reduces noise while avoiding blur caused by pre- or post-denoising. The method comprises two deep sub-networks: an Image Contrast Enhancement Network (ICE-Net) and a Re-Enhancement and Denoising Network (RED-Net). The ICE-Net takes a low-light image as input and produces a contrast-enhanced image. The RED-Net takes both the output of ICE-Net and the original low-light image as input, enabling it to re-enhance the low-light image and denoise simultaneously. Both networks can be trained using only low-light images, achieved through a Maximum Entropy based Retinex (ME-Retinex) model and an assumption that noise is independently distributed. In the ME-Retinex model, we introduce a novel constraint on the reflectance image: the maximum channel of the reflectance image should conform to the maximum channel of the low-light image while having maximum entropy. This converts the decomposition of reflectance and illumination in the Retinex model from an ill-conditioned to a well-posed problem, enabling self-supervised training of ICE-Net.

The loss functions of RED-Net are carefully formulated to separate noise from details during training, based on the principle that if noise is independently distributed, the gradient of the noise component should be smaller than that of the detail component after processing with smoothing filters (e.g., mean filters). The effectiveness of the proposed method is demonstrated both qualitatively and quantitatively through experiments.

1. Introduction

Images captured in low-light conditions invariably suffer from low contrast, low brightness, and severe noise. Low-light image enhancement methods address these problems before high-level computer vision tasks, yet few methods handle these issues effectively simultaneously. Recently, various deep learning-based algorithms have achieved remarkable results in image processing and computer vision tasks such as object detection [?, ?, ?, ?] and image segmentation [?, ?, ?]. A key reason for this rapid development is the availability of large datasets with clear and unambiguous labels. Although dataset construction incurs some cost, it remains acceptable, and numerous open-source datasets can be found online to support network training.

However, in low-level image processing tasks such as low-light image enhancement, image dehazing, and image restoration, obtaining large numbers of true input-label image pairs is difficult. For low-light image enhancement, previous supervised solutions have achieved good visual effects, particularly in noise reduction, by synthesizing low-light images [?] or using images with different exposure times [?]. Nevertheless, two fundamental problems remain. First, it is unclear how to ensure that a pre-trained network generalizes to images from different devices, scenes, and lighting conditions without building new training datasets (e.g., [?] fails to remove background noise in Fig. 1). Second, it is difficult to determine whether the normal-light image used for supervision is optimal, as multiple normal-light images may correspond to a single low-light image. Typically, datasets with paired low/high-light images require artificial adjustment, which is time-consuming and labor-intensive, and we cannot guarantee that the normal-light images are ideal for training.

To overcome these limitations, we propose a self-supervised low-light image enhancement framework that simultaneously performs image contrast enhancement and denoising. Similar to previous works [?], we use two networks for contrast enhancement and denoising, respectively. However, unlike prior work, both the Image Contrast Enhancement Network (ICE-Net) and the Re-Enhancement and Denoising Network (RED-Net) are trained in a self-supervised manner. The RED-Net is specifically designed to reduce noise by re-enhancing the contrast of low-light images, thereby minimizing information loss caused by pre- or post-processing with existing denoising methods.

For training ICE-Net, we propose a Maximum Entropy based Retinex (ME-Retinex) model. Unlike previous Retinex models that only assume smooth illumination, our ME-Retinex model introduces a new constraint on the reflectance image: the maximum channel of the reflectance image should conform to the maximum channel of the low-light image while having maximum entropy. This constraint allows direct control over the image enhancement level and converts the ill-posed decomposition of reflectance and illumination into a well-posed problem.

For training RED-Net, we adopt the assumption that noise follows a Poisson distribution and is independent across different pixels. Under this assumption, the gradients of noise should be smaller than those of details after smoothing filter processing, with most noise gradients approaching zero. This enables separation of noise from details in reflectance by treating gradients from smoothed reflectance as weights. Since our task involves image contrast enhancement, where edges with higher gradients typically have higher contrast, the RED-Net loss functions are designed to increase gradients of details and edges for better contrast enhancement—differing from previous works that only preserve edges. Based on these ideas, we formulate several self-supervised loss functions and validate their effectiveness through experiments.

The loss functions in this paper enable complete self-supervised training, meaning we can solve image enhancement tasks for specific images using either CNNs or analytical methods. However, more training data in CNNs often yields better results (e.g., Fig. 4), and CNNs typically require less processing time than analytical methods. The proposed method is independent of how low-light images are acquired, and the training process is entirely self-supervised, providing good generalization capability. Even if the pre-trained network performs poorly in a new environment, retraining or fine-tuning is possible without constructing paired or unpaired normal-light image datasets. Our contributions are summarized as follows:

- We propose a framework for enhancing low-light images that simultaneously improves image contrast and reduces noise. Through close coupling of the two processes, we reduce information loss in image enhancement tasks (e.g., blur caused by commonly used post-denoising).
- We propose an Image Contrast Enhancement Network (ICE-Net) and a Re-Enhancement and Denoising Network (RED-Net), both trainable via self-supervision, eliminating dependence on paired or unpaired images. The RED-Net can be combined with other Retinex or HSV-based image enhancement methods to achieve re-enhancement and noise suppression, even for methods like AHE (Adaptive Histogram Equalization [?]) that produce heavy noise, which benefits many previous contrast enhancement studies.
- We compare the proposed method with several state-of-the-art methods through comprehensive experiments, evaluating results using objective metrics and visual quality. All results consistently demonstrate the effectiveness of our method.

2. Related Works

Low-light image enhancement. Directly adjusting image contrast is perhaps the most intuitive approach to enhancement, including Histogram Equalization (HE) and other HE-based improvements [?, ?, ?, ?, ?]. Although these improved methods aim to suppress noise, preserve hue, and maintain brightness, direct contrast adjustment still suffers from over- and under-enhancement, noise amplification, and other issues. Gamma correction is another mapping approach frequently used for low-light enhancement. While it ensures good image brightness and stretches contrast in dark or bright areas, it cannot avoid noise amplification, and its results heavily depend on the manually selected Gamma value.

Retinex is a widely used model for low-light enhancement in recent years. According to Retinex theory, an image can be decomposed into reflectance and illumination. Early works SSR [?] and MSR [?] treat reflectance as the final enhanced result. However, since decomposition is an ill-posed problem without sufficient constraints on reflectance, enhanced images often exhibit unrealistic phenomena such as over-enhancement and whitening, and these methods struggle with noise reduction. Recent works enhance illumination after decomposition and obtain the final enhanced image by recombining enhanced illumination and reflectance. However, enhanced images may still contain noise, requiring additional post-denoising procedures [?, ?] that produce blurred details. [?] introduced a joint low-light enhancement and denoising method that achieves both simultaneously, further improving the method by considering a noise map compared to conventional Retinex models. Although these methods show promising results, most require multiple iterations for decomposition, which is time-consuming. Meanwhile, without automatic illumination manipulation, enhanced images may lack proper contrast and typically require careful parameter tuning.

Recently, the remarkable performance of deep learning has inspired promising works in low-light image enhancement, including supervised methods [?, ?, ?, ?] and unsupervised methods [?, ?]. Most early supervised learning works train networks with synthetic datasets [?, ?, ?, ?]. Although the data obtained by these methods appears dark and noisy, it differs from natural low-light images. Chen et al. [?] introduced a dataset containing real raw low-light images and corresponding raw high-light images for training. Since multiple reference images may exist for a single input low-light image, they introduced an amplification ratio in the network to achieve correspondence between input and reference. This method effectively solves noise and color distortion problems; however, the ratio must be manually chosen during testing, limiting widespread use, and a single ratio may produce over- or under-enhanced areas. [?] introduced the LOL dataset containing real paired low and high-light images, incorporating the Retinex model into training to connect reflection images of input and reference, and proposed denoising reflectance with BM3D [?]. However, this approach still causes blur or leaves noise. [?] added a subnet called restoration-net to denoise reflectance and provided an extra brightness ratio to control illumination, but testing still requires manual ratio adjustment for better results. Although these methods use real low-light data for training, the lack of constraints on enhanced image contrast leads to over-enhancement (saturation) or under-enhancement, even with artificial parameter adjustment. In unsupervised works, [?] proposed a GAN-based method trainable with unpaired data but cannot control enhancement results. [?] proposed a zero-reference low-light enhancement method trainable without any paired or unpaired data but did not provide noise removal methods.

Image denoising. Many denoising methods have been proposed over the past few decades, including conventional methods [?, ?] and learning-based methods [?, ?, ?]. However, these denoising methods are not specifically designed for low-light image enhancement tasks. Pre- or post-processing with these methods causes detail loss, and learning-based methods may become invalid for different noise distributions. [?] proposed a denoising method for low-light enhancement but requires paired low/high-light image data for training. Recently, [?] proposed an unsupervised denoising method called N2V that can be trained with only noisy images; however, in our tests, it still causes blur even when retrained with enhanced images.

[Figure 2: see original paper]. The structure of the proposed method. RED-Net takes low-light images and the max channel of the output of ICE-Net as input.

3. Proposed Model

The proposed method aims to achieve low-light image enhancement and denoising without any artificial adjustment when only low-light images are available. For example, when a camera enters a new low-light environment, pre-trained networks may not work for different distributions, and the only available data are low-light images. To achieve automatic contrast enhancement, we propose a Maximum Entropy based Retinex model and a self-supervised ICE-Net that leverages multiple images. For denoising, we propose a self-supervised RED-Net specifically designed for low-light image enhancement tasks. Through the combination of re-enhancement and denoising, RED-Net can preserve more details during the denoising process. The structure of the proposed method is shown in Fig. 2.

3.1. ME-Retinex Model and ICE-Net

Recently, many low-light image enhancement works are based on the following Retinex model:

$$S = R \circ I$$

where S and I represent the captured image and illumination image, respectively, R represents reflectance (treated as the desired enhanced image in some works), and $\circ$ denotes element-wise multiplication. Most recent works assume the three color channels share the same illumination to simplify the model [**?**, **?**], generally using the maximum value of the three color channels as the initial illumination map estimate [**?**]. It has been proven that image enhancement methods based on this simplified Retinex model are equivalent to directly controlling contrast on the V channel in HSV color space while keeping H and S channels unchanged [**?**]. However, differences remain between these two approaches. Methods that directly control image contrast typically stretch contrast in some areas while compressing it in others, causing detail loss in compressed areas (over-/under-enhancement can be considered detail loss). For example, HE merges smaller bins, and Gamma correction compresses contrast in bright areas, both causing detail loss. Retinex-based methods usually lack constraints on the contrast of the target enhanced image (whether R or $R \circ I^{\gamma}$), producing uncertain results.

If we interpret HE or Gamma correction through the Retinex model, the enhanced image R is obtained via S/I without the constraint that illumination I is smooth, so missing details are retained in illumination I. Consider that if a rich texture area S is divided by a smooth I, details will appear in R, avoiding detail loss in HE or Gamma correction. We can thus combine direct contrast control methods with the Retinex model to leverage both advantages. This combination also suppresses noise through the assumption that illumination is smooth, compared to direct contrast control methods.

Typically, a Retinex-based method can be expressed as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \lambda_1 \mathcal{L}_R + \lambda_2 \mathcal{L}_I$$

where $\mathcal{L}_{recon}$, $\mathcal{L}_R$, and $\mathcal{L}_I$ represent reconstruction loss, reflectance loss, and illumination loss, respectively, and λ_1 and λ_2 are weight parameters. The reconstruction loss $\mathcal{L}_{recon}$ can be expressed as:

$$\mathcal{L}_{recon} = \|S - R \circ I\|_1$$

where $\|\cdot\|_1$ denotes the L1 norm. We use the L1 norm to constrain all losses and do not compare the impact of L1, L2, SSIM, and other loss functions on low-level image processing tasks, as related studies already exist [**?**].

In this paper, we adopt the HE method to form a Maximum Entropy based Retinex model, formulating the reflectance loss as:

$$\mathcal{L}_R = \left\| \max_{c \in \{r,g,b\}} R_c - \mathcal{F}\left(\max_{c \in \{r,g,b\}} S_c \right) \right\|_1 + \lambda \|\nabla R\|_1$$

where $\mathcal{F}(x)$ denotes the histogram equalization operator applied to image x, λ is a weight parameter, and ∇ denotes the gradient operator. The first term means the maximum channel of reflectance should conform to the maximum channel of the low-light image with maximum entropy, which can be considered as directly controlling the contrast of the enhanced image. The second term is a commonly used smoothing term to suppress noise, though it is often difficult to distinguish image details from noise effectively.

For illumination loss, we adopt the structure-aware smoothness loss proposed in [**?**]:

$$\mathcal{L}_I = \|\nabla I \circ \exp(-\lambda_3 |\nabla R|)\|_1$$

Equation 5 is proposed to make the illumination loss aware of image structure [**?**]. This loss means the original TV function $\|\nabla I\|_1$ is weighted by the gradient of reflectance.

For Equations 2-5, we introduce an ICE-Net to solve this optimization problem. One might ask: why introduce a CNN? Since the ideal image can be obtained by minimizing the total loss, one could simply run this optimization directly on R and I for a single image S, making the CNN seem unnecessary. However, most optimization processes require multiple iterations, causing time consumption issues, and more constraints lead to more complex solutions. Additionally, HE is a global enhancement method that inevitably causes locally over-bright or over-dark problems. By introducing CNN and training on multiple images, this problem can be avoided, as shown in Fig. 4 [Figure 4: see original paper]. This is because, under the HE constraint, the same local area in different images will be enhanced to different degrees, and CNN trained with L1 regularization will learn to find median values instead of becoming over-bright or over-dark.

Furthermore, the loss function and ICE-Net are designed to learn appropriate contrast and brightness enhancement, so we made no special design for denoising; the RED-Net designed in the next subsection can effectively achieve denoising.

3.2. RED-Net

After ICE-Net processing, although image contrast has improved, some noise remains. Inspired by [?], which introduced a Conditional Re-Enhancement Network (CRE-Net) for denoising in low-light enhancement tasks, we further propose a self-supervised RED-Net to re-enhance low-light images and denoise simultaneously.

In this part, we still build the loss function based on Equation 2, but each sub-loss function has been modified. For the reconstruction loss $\mathcal{L}'_{recon}$ in RED-Net, we adopt the assumption that noise follows a Poisson distribution [?], which better matches real low-light image noise. To distinguish variables in RED-Net from those in ICE-Net, we add the superscript $(\cdot)'$ for RED-Net variables. The reconstruction loss can be expressed as:

$$\mathcal{L}'_{recon} = R' \circ I' - S \circ \log(R' \circ I')$$

where R' and I' represent the reflectance and illumination produced by RED-Net, respectively, and R' is also the target enhanced image of the entire proposed method.

We argue that an image can be divided into different components: noise, flat areas, details, and structural information, with no clear dividing line between details and structural information. Many methods can remove noise, but the key challenge is separating details and structure from noise, then preserving or even strengthening those details and structure during denoising. To make reflectance less noisy while preserving rich details and sharp edges, we design the reflectance loss as follows:

$$\mathcal{L}'_R = \max_{c\in\{r,g,b\}} R'_c - \max_{c\in\{r,g,b\}} R_c \circ \log\left(\max_{c\in\{r,g,b\}} R'_c\right) + \lambda \|W \circ \mathcal{N}(|\nabla R'|) \circ \exp(-\lambda_3 W \circ \mathcal{N}(|\nabla R'|))\|_1$$

where $\mathcal{N}(x)$ and $|x|$ represent local normalization on x and absolute value of x, respectively. R and R' represent the output reflectance of ICE-Net and RED-Net, respectively. W represents weights calculated as:

$$W = \mathcal{N}(|\nabla(\mathcal{G}(R'))|)$$

where $\mathcal{G}(x)$ denotes a smoothing filter on x (mean filter is used in the proposed method). The curve of the second term $x \cdot \exp(-\lambda x)$ is shown in Fig. 3 [Figure 3: see original paper]. Intuitively, after smoothing, details and structure still have gradients, though smaller than before, while noise and smooth areas have no gradients or much smaller gradients. We can thus use gradients from smoothed images as weights. As shown in Fig. 3, when the loss function takes the form $x \cdot \exp(-\lambda x)$, small x becomes smaller and high x becomes higher during training. Through local normalization, details and structure are more likely to fall on the right side, making them sharper during training. As shown in Fig. 5 [Figure 5: see original paper], noise is well removed while details are preserved.

[Figure 3: see original paper]. The curve of $y = x \cdot \exp(-\lambda x)$

[Figure 4: see original paper]. Results by ICE-Net with different training data. (a) Input. (b) Network trained with multiple data. SSIM: 0.6743, PSNR: 23.4716, NIQE: 3.9140 (c) Network trained with (a) only. SSIM: 0.4858, PSNR: 15.5112, NIQE: 4.9367 (d) Reference

Typically, illumination is expected to retain only structural information while ignoring detailed information. Therefore, we can adopt a design similar to the reflectance loss, and the illumination loss can be expressed as:

$$\mathcal{L}'_I = \|W_I \circ \mathcal{N}(|\nabla I'|) \circ \exp(-\lambda_3 W_I \circ \mathcal{N}(|\nabla I'|)) \circ \exp(-\lambda_3 W_R \circ \mathcal{N}(|\nabla R'|))\|_1$$

where W_I and W_R represent weights calculated as:

$$W_I = \mathcal{N}(|\mathcal{G}(\nabla I')|)$$
$$W_R = \mathcal{N}(|\mathcal{G}(\nabla R')|)$$

where $\mathcal{G}(x)$ and $\mathcal{N}(x)$ still denote smoothing filter and local normalization on x, respectively. Unlike W in Equation 8, the order of gradient operation ∇ and smoothing operation $\mathcal{G}$ are switched.

It can be considered that for noise and details, the mean gradient in a local area should be small, which is quite different for structure. For example, text on white paper may have opposite gradients in a local area, making the mean gradient close to zero. With W_I and the special loss form $x \cdot \exp(-\lambda x)$, we can separate noise and details from structure in illumination and make structural edges sharper during training. We also preserve $\exp(-\lambda_3 W_R \circ \mathcal{N}(|\nabla R'|))$ and introduce weight W_R to ensure consistency of structural information between reflectance and illumination. It should be noted that all weight terms W, W_R, and W_I do not participate in backpropagation during training.

4. Experiments

We use the LOL database [?], which contains 500 low/normal-light image pairs, with 485 used for training. Each image size is 400×600. Note that during training, we only use natural low-light images without any synthetic data or normal-light images. During training, our batch size is set to 16 and patch size to 48×48. We use Adam stochastic optimization [?] to train the network with a learning rate of 0.001. Network training and testing are completed on an Nvidia GTX 2080Ti GPU and Intel Core i9-9900K CPU, with code based on the TensorFlow framework.

To evaluate the proposed method's performance on enhancing low-light images, we quantitatively and visually compare it with several low-light enhancement methods, including LIME [?], RRM [?], Retinex-Net [?], KinD [?], and additional data collected from other datasets for testing. Three metrics are adopted for quantitative comparison: Peak Signal-to-Noise Ratio (PSNR), Structural SIMilarity (SSIM) [?], and NIQE [?]. NIQE is a non-reference image quality assessment method that evaluates image naturalness, with lower values indicating better quality. PSNR and SSIM are reference-based quality assessment methods indicating noise level and structural similarity between results and reference, respectively.

4.1. Ablation Study

To demonstrate the necessity of introducing CNN and the effectiveness of each component, we conduct two ablation studies.

Contribution of ICE-Net. This ablation study addresses why we do not simply optimize the loss function to obtain results, as in other variational-based Retinex models [?, ?], if the network can be trained in a self-supervised way. As mentioned in Sec. 3.1, the CNN-based ICE-Net is introduced to avoid HE-related problems through training with multiple data. Since directly solving Equation 2 through variational methods under our proposed loss functions is difficult, we use a CNN trained with only a single low-light image instead, with results considered a solution to Equation 2.

Fig. 4 presents ICE-Net results trained with a single low-light image versus multiple images, showing that training with multiple data achieves better contrast and brightness enhancement. As seen in Fig. 4(c), optimization on a single low-light image cannot avoid under- or over-enhancement problems (e.g., green pipes and metal hinges) caused by HE. However, in Fig. 4(b), training with multiple images produces more appropriate brightness in every local area. Objective metrics also show better results with multiple training data in PSNR, SSIM, and NIQE, indicating less noise and results more similar to the reference and more natural.

Contribution of Each loss in RED-Net. We present RED-Net results trained with different losses and weights in Fig. 5. To better illustrate the importance of each loss component, we use AHE [?], which produces serious noise during processing, as the contrast enhancement method (e.g., Fig. 5(c)), and study re-enhancement and denoising effects under different loss functions. We use the complete loss function as the baseline (containing Equations 6-11) and study the influence of removing different weight terms:

- Without W in Equation 7 (Fig. 5(e))
- Without W_I and W_R in Equation 9 (Fig. 5(f))
- Without $\exp(-\lambda_3 W_R \circ \mathcal{N}(|\nabla R'|))$ in Equation 9 (Fig. 5(h))
- Without $\exp(-\lambda_3 W_I \circ \mathcal{N}(|\nabla I'|))$ in Equation 9 (Fig. 5(g))

As shown in Fig. 5(d), with all proposed loss functions, RED-Net can obviously reduce noise while preserving details. When we simply remove W (Fig. 5(e)), only obvious structures are preserved, proving the importance and effectiveness of separating noise and details through W. When we remove $\exp(-\lambda_3 W_R \circ \mathcal{N}(|\nabla R'|))$ (Fig. 5(h)), details are lost and some obvious edges become slightly blurred. When we remove W_I and W_R, designed to smooth noise and details while preserving structure in illumination, some details in reflectance are blurred (Fig. 5(f)), proving the importance of smoothing noise and details in illumination and validating our design.

In Fig. 5(g) and (d), it appears that removing $\exp(-\lambda_3 W_I \circ \mathcal{N}(|\nabla I'|))$ from illumination loss does not affect reflectance results. However, as seen in Fig. 6 [Figure 6: see original paper], edges (e.g., edges in red rectangles) in illumination are blurred under this loss variant, which may cause halo effects in reflectance. Well-behaved illumination can also help future work (e.g., avoiding over-enhancement). We also studied the case without Poisson distribution; however, with AHE [?], RED-Net output becomes completely unacceptable, even losing structure.

[Figure 6: see original paper]. Visual comparison with different training losses in RED-Net. (a)-(b) With or without $\exp(-\lambda_3 W_I \circ \mathcal{N}(|\nabla I'|))$, see details respectively. Please zoom in to view.

4.2. Comparison with State-of-the-Arts

This subsection compares the proposed method's performance with current state-of-the-art methods through qualitative and quantitative experiments. We used not only the LOL dataset but also standard datasets from previous works, including LIME [?] (10 images), MF [?] (10 images), and VV [?] (23 images).

We compared the combination of ICE-Net and RED-Net with previous methods capable of contrast enhancement and denoising, including LIME [?] with denoising post-processing, RRM [?] for joint enhancement and denoising, Retinex-Net [?] trained supervisedly and denoised with BM3D [?] in reflectance, and KinD [?] trained supervisedly for contrast enhancement and denoising. Code was downloaded from authors' homepages and parameters set as recommended. Results are shown in Fig. 7 [Figure 7: see original paper] and Tables 1 and 2 .

Fig. 7 shows qualitative evaluation results. Compared with LIME and Retinex-Net, which denoise reflectance with BM3D, our two-stage method that first enhances then re-enhances and denoises achieves a better balance between denoising and detail preservation, even comparable to the supervised method KinD (e.g., books in the bookcase processed by LIME are blurred, Retinex-Net still has serious noise, while both our method and KinD preserve text well). Also, as seen in the second and last rows of Fig. 7, our method works under serious noise and non-uniform illumination conditions (e.g., face and arm in shadow are well enhanced). Since we assume the difference between detail and noise is independent noise distribution—which is not always true—the rough wall was smoothed in the blue rectangle in the third row. (More detailed experiments, comparisons, network structures, and parameters are included in supplementary material. In our experiments, network structure impact is not significant; RED-Net has the same structure as in [?], and ICE-Net is similar to RED-Net but with fewer layers.)

Tables 1 and 2 show quantitative evaluation results. Our method achieves poorer NIQE but highest PSNR and middle SSIM, meaning our processed images differ from natural images and references after final enhancement, but noise is well removed. This is because, when designing ICE-Net and RED-Net, we mainly considered automatic adaptation to new environments and noise removal, without incorporating natural image priors into loss functions, especially in RED-Net. Meanwhile, during enhancement, our goal is to enhance details in each local area, which differs from reference images obtained by adjusting exposure time.

. NIQE scores on each subset (LOL [?], LIME [?], MF [?], VV), where smaller NIQE indicates better alignment with natural images.

Dataset	LIME [?]	LOL [?]	MF [?]	VV
LIME[?]				
RRM[?]				
Retinex-Net[?]				
KinD[?]				
Proposed				

. SSIM and PSNR scores on the LOL [?] dataset, where higher SSIM and PSNR indicate better alignment with reference and less noise, respectively.

Dataset	LIME[?]	RRM[?]	Retinex-Net[?]	KinD[?]	Proposed
SSIM					

Dataset	LIME[?]	RRM[?]	Retinex-Net[?]	KinD[?]	Proposed
PSNR					

5. Conclusion and Future Work

This paper proposes a two-stage framework for automatic low-light image enhancement and denoising that first enhances image contrast then further re-enhances and denoises. Both networks in our method can be trained in a self-supervised manner, enabling application in real new environments and devices. Experimental results on various low-light datasets show our method is comparable to many state-of-the-art methods in both visual effects and objective metrics. Future work will explore restoring color degradation, combining RED-Net and ICE-Net, and integrating low-light image enhancement with high-level tasks to further improve real-time performance.

References

[1] F. Agostinelli, M. R. Anderson, and H. Lee. Adaptive multi-column deep neural networks with application to robust image denoising. *Advances in Neural Information Processing Systems*, pages 1493–1501, 2013. 3

[2] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017. 1

[3] Turgay Celik and Tardi Tjahjadi. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, 20(12):3431–3441, 2011. 2

[4] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018. 1, 3

[5] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 3, 7

[6] Qingxu Fu, Xiaoguang Di, and Yu Zhang. Learning an adaptive model for extreme low-light raw image processing. *arXiv preprint arXiv:2004.10447*, 2020. 3

[7] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 129:82–96, 2016. 8

[8] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2782–2790, 2016. 3

[9] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. In *2014 IEEE Conference on Computer*

Vision and Pattern Recognition, pages 2862–2869, 2014. 3
[10] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong. Zero-reference deep curve estimation for low-light image enhancement. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1777–1786, 2020. 3
[11] 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, 2016. 3, 4, 6, 7, 8
[12] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 1
[13] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *arXiv preprint arXiv:1906.06972*, 2019. 3
[14] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997. 3
[15] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462, 1997. 3
[16] Ron Kimmel, Michael Elad, Doron Shaked, Renato Keshet, and Irwin Sobel. A variational framework for retinex. *International Journal of computer vision*, 52(1):7–23, 2003. 6
[17] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6
[18] A. Krull, T. Buchholz, and F. Jug. Noise2void - learning denoising from single noisy images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2124–2132, 2019. 3
[19] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22(12):5372–5384, 2013. 2
[20] Chongyi Li, Jichang Guo, Fatih Porikli, and Yanwei Pang. Lightennet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognition Letters*, 104:15–22, 2018. 3
[21] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 3, 6
[22] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 7, 8
[23] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S. Huang. Non-local recurrent network for image restoration. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 1680–1689, Red Hook, NY, USA, 2018. Curran Associates Inc. 3
[24] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott

Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016. 1

[25] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 1

[26] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Ll-net: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2015. 1, 2, 3

[27] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2012. 6

[28] Sarif Kumar Naik and CA Murthy. Hue-preserving color image enhancement without gamut problem. *IEEE Transactions on Image Processing*, 12(12):1591–1598, 2003. 2

[29] Seonhee Park, Soohwan Yu, Byeongho Moon, Seungyong Ko, and Joonki Paik. Low-light image enhancement using variational optimization-based retinex model. *IEEE Transactions on Consumer Electronics*, 63(2):178–184, 2017. 6

[30] Etta D Pisano, Shuquan Zong, Bradley M Hemminger, Marla DeLuca, R Eugene Johnston, Keith Muller, M Patricia Braeuning, and Stephen M Pizer. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital imaging*, 11(4):193, 1998. 2

[31] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987. 2, 6, 7

[32] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 1

[33] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015. 1

[34] X. Ren, M. Li, W. Cheng, and J. Liu. Joint enhancement and denoising method via sequential decomposition. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2018. 3

[35] Liang Shen, Zihan Yue, Fan Feng, Quan Chen, Shihao Liu, and Jie Ma. Msr-net: Low-light image enhancement using deep convolutional network. *arXiv preprint arXiv:1711.02488*, 2017. 3

[36] Yu Wang, Qian Chen, and Baeomin Zhang. Image enhancement based on equal area dualistic sub-image histogram equalization method. *IEEE Transactions on Consumer Electronics*, 45(1):68–75, 1999. 2

[37] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 3, 4, 6, 7, 8

[38] Jie Yang, Xinwei Jiang, Chunhong Pan, and Cheng-Lin Liu. Enhancement

of low light level images with coupled dictionary learning. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 751–756. IEEE, 2016. 3

[39] Q. Yang, C. Jung, Q. Fu, and H. Song. Low light image denoising based on poisson noise model and weighted tv regularization. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3199–3203, 2018. 5

[40] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 3

[41] Yu Zhang, Xiaoguang Di, Bin Zhang, Ruihang Ji, and Chunhui Wang. Better than reference in low light image enhancement: Conditional re-enhancement networks. *arXiv preprint arXiv:2008.11434*, 2020. 3, 4, 5, 8

[42] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1632–1640, 2019. 1, 2, 3, 4, 6, 7, 8

[43] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016. 4

[44] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6

Note: Figure translations are in progress. See original paper for figures.