MSIGEN: Multi-Scale Illumination-Guided Low-Light Image Enhancement Network

Yue Hao¹, Xuemei Jiang², Jiwei Hu¹* and Ping Lou¹

¹ Information and Communication Engineering, Wuhan University of Technology, Wuhan, Hubei, 430070, China
² Information and Communication Engineering, Wuhan University of Technology, Wuhan, Hubei, 430070, China
*hujiwei@whut.edu.cn

Abstract. Images lack of light often show the characteristics of low visibility, low contrast and high noise, due to limited imaging equipment. To address the problem, inspired by Retinex theory, we propose a two-stage method called Multi-Scale Illumination-Guided Low-Light Image Enhancement Network (denoted as MSIGEN). In the first stage, we employ an enhancement module to achieve low-light image enhancement, including a Decom-Net for decomposing into illumination and reflectance, an Adjust-Net for illumination adjustment, a Restore-Net for the illumination fusion map guided reflectance restoration. Secondly, the enhancement module is introduced to refine the initial images to effectively remove visual defects amplified and ensure the naturalness of the image. In addition, extensive experiment results demonstrate the advantages of our method and the effect of this method has reached the state-of-the-art.

1. Introduction
Images which are shot in low light conditions, usually contain low visibility, low contrast, artifacts, unexpected noise defects. Low-light images lose details, which lead to many problems, such as unpleasant feelings, the poor performance of many computer vision systems. So low-light image enhancement is an arduous and important task to handle many factors simultaneously and effectively.

In the past few years, although some equipment that can capture in low light conditions has been developed, the technology is not very mature and cannot be widely used. It is also worth noting that the core technology of these devices is still a low-light image enhancement algorithm.

The pioneering categories are histogram equalization [1] and Retinex theory [2]. The former increases the image contrast but it doesn’t consider the global contrast in the entire image. Built on the latter, Single-scale Retinex (SSR) [3] and Multi-scale Retinex called MSRCR [4] constrain the light map to be smooth through a Gaussian filter, but the outputs usually look unnatural. To solve the problem of unnaturalness, a new method called NPE was proposed by Wang et al. [5] that can enhance the contrast and keep the illumination natural at the same time. Fu et al. proposed the SRIE method [6], using a weighted variation model to reckon the illumination and reflectance at the same time. Lore et al. [7] designed a method for low-light image enhancement (LLNet) based on deep self-encoding that can simultaneously achieve contrast enhancement and denoising. Then a deep network denoted as Retinex-Net [8] is designed by Wei et al, which achieves the Retinex theory in a deep learning way, and uses a
ready-made denoising tool (BM3D [9]), but the different effects of noise are not good. Zhang et al. [10] proposed a network (KinD) which achieve the transform of different light/exposure conditions.

Therefore, in this article, we propose a two-stage enhancement method, which is an adaptive illumination compensation method, to achieve low-light image enhancement and color naturalness and denoising in the process of enhancement. On the basis of learning Retinex theory, in the decomposition step, the input image is decomposed into the reflectance and illumination. The illumination adjustment achieves to flexibly convert one light condition to another one. In the reflectance restoration step, multi-scale illumination is introduced to adjust the reflectance. On the basis, a reinforce module is proposed to fine-tune the original enhanced image, which achieves denoising simultaneously and effectively to restore the image color.

In general, our contribution is reflected in three aspects:
1) We propose a two-stage method and a multi-scale illumination fusion network architecture to achieve reflection restoration. Guided by the illumination map, the proposed approach can achieve denoising simultaneously and effectively to restore the image color.
2) In our network, there is also a reinforce module, which can fine-tune the image to ensure the naturalness of the image and effectively remove visual defects amplified.
3) Extensive experimental results demonstrate the advantages of our method and our method has reached the state-of-the-art in terms of effect.

Figure 1. Demonstration of our framework (MSIGEN). The proposed MSIGEN consists of two stages: (a) enhancement and (b) reinforce. The enhancement module includes three main steps: decomposition, illumination adjustment and reflectance restoration. Then, we propose a reinforce module to fine-tune the enhanced images, which is capable to validly decrease the visual unnaturalness and fine-tunes the image to ensure the naturalness of the image.

2. Proposed Method
Base on Retinex model, an image can be composed of the reflectance and illumination map as follows:

$$S = R \odot I$$  \hspace{1cm} (1)

where $S$ represents the captured image, $R$ represents the reflectance and $I$ represents the illumination map, $\odot$ represents element multiplication. Inspired by Retinex theory, we design a two-stage method called Multi-Scale Illumination-Guided Low-Light Image Enhancement Network (denoted as MSIGEN). As illustrated in Figure 1, our MSIGEN approach is made up of two components: enhancement and reinforce.

Stage-I: Enhancement
The enhancement module is introduced to achieve low-light image enhancement. As shown in Figure 1 (a), the enhancement module includes three main parts: decomposition, illumination adjustment and
reflectance restoration. In the first stage, we propose a decomposition module to decompose the input into illumination and reflectance, an adjustment module to flexibly convert one light condition to another one, a restoration module to solve the noise issues for the attention map guided reflectance restoration. Then, the adjusted illumination and reflectance are reproduced to the initial image.

**Decomposition.** This is an uncertain solution problem. Its solution needs additional prior, and the loss of image enhancement can be disintegrated into the form of four distance terms, as can be seen in formula (2):

\[
L = L_{\text{rec}} + L_{\text{rs}} + L_{\text{is}} + L_{\text{mc}}
\]

where, \(L_{\text{rec}}\) represents reconstruction loss, \(L_{\text{rs}}\) represents reflectance loss, and \(L_{\text{is}}\) represents illumination loss, \(L_{\text{mc}}\) represents mutual consistency.

In this paper, we should combine the illumination and reflectance to be the input, which is defined by the reconstruction loss \(L_{\text{rec}}\). We use the \(L_1\) norm to constrain the reconstruction loss, as can be seen in formula (3):

\[
L_{\text{rec}} = ||I_1 - R_1 \circ L_1||_1 + ||L_1 - R_1 \circ L_1||_1
\]

Based on Retinex theory, we regularize the decomposed reflectance pair to be close. So the reflectance similarity is constrained by:

\[
L_{\text{rs}} = ||R_1 - R_1||_2^2
\]

The illumination smoothness loss is constrained by:

\[
L_{\text{is}} = \left\| \frac{\nabla L_1}{\max(|\nabla L_1|,0.01)} \right\|_1 + \left\| \frac{\nabla L_1}{\max(|\nabla L_1|,0.01)} \right\|_1
\]

where \(\nabla\) denotes the gradient including \(\nabla_x\) (horizontal) and \(\nabla_y\) (vertical) directions.

We can preserve strong correlation edge information and depress weak correlation edge information through the mutual consistency loss. The loss well fits the mutual consistency, and the aim is to preserve edge information, which is constrained by:

\[
L_{\text{mc}} = ||M \cdot \exp(-c \cdot M)||_1
\]

where \(M = |\nabla L_1| + |\nabla L_1|\).

**Illumination adjustment.** In this step, the enhancement module calculates the paired illuminations’ ratio of strength, i.e. \(\alpha = L_1/L_1\). The ratio of strength \(\alpha\) is defined to make the illumination condition from a low light \(L_1\) to a target one \(L_1\). Considering that the output image \(\hat{L}\) should be similar to the target image and the similarity between their edges, the following is the loss, as can be seen in formula (7):

\[
L_{\text{lt}} = ||\hat{L} - L_1||_2^2 + ||\nabla \hat{L}||_1^2 + ||\nabla L_1||_2^2
\]

where \(||,||_2\) means \(L_2\) norm.

**Reflectance restoration.** Firstly, each image is processed by a U-net branch with the same architecture. We add skip connections in the U-net to aid the reconstruction of details at different scales. Secondly, we propose fusion blocks to combine the illumination map to fully exploit valuable information in a complementary way (see Figure 2). The fusion blocks are build up a permutation-invariant technique with more aggregation operations between features. The input of the fusion blocks is the output of each layer and illumination map. Each fusion block takes different kernel sizes to combine illumination features into each layer. In this part, due to the reflectance from strong-light image is less disturbed than the low. We use a clearer reflectance as ground truth that is the reflectance from strong-light image. The following is the restoration function loss, as follows:

\[
L_{\text{R}} = ||\hat{R} - R_1||_2^2 + 1 - \text{SSIM}(\hat{R}, R_1) + ||\nabla \hat{R} - \nabla R_1||_2^2
\]

where SSIM (\(\cdot,\cdot\)) is the structural similarity measurement, and \(\hat{R}\) corresponds to the output of the module( stored reflectance).
Stage-II: Reinforce

The motivation is to refine the initial images to validly change the visual unnaturalness and improve the details. As shown in Figure 1 (a), the Reinforce module includes two subnets: an Estimation-Net and a Reinforce-Net. The Estimation-Net is designed to estimate the noise map \(\hat{x}\) which guides the method. The Reinforce-Net is designed to contrast re-enhancement to remove visual defects amplified and change the visual unnaturalness. We use the output of reflectance restoration \(x\) as the input. We directly adopt FCN in our implementation in the Estimation-Net. Each Conv layer deploys the ReLU nonlinearity [12] except the last one. As for Reinforce-Net, we adopt a U-Net [11] architecture, which takes both the \(x\) and the \(\sigma(x)\) as input to give a prediction of the normal image. We adopt the skip structure to connect features, convolutions with different strides and transpose convolutions to exploit multi-scale information.

In this module, we propose a loss function to fine-tune the image, which is expressed as:

\[
L = \|\hat{x} - x\|^2 + \lambda_{TV} \|\nabla_x \sigma(x)\|^2 + \|\nabla_y \sigma(x)\|^2
\]

(9)

Where \(\|\hat{x} - x\|^2\) means the reconstruction loss, \(\lambda_{TV}\) denote the TV regularizer, \(\|\nabla_x \sigma(x)\|^2\) and \(\|\nabla_y \sigma(x)\|^2\) denote that \(\sigma(x)\) becomes more smooth.

3. Experiment

3.1. Implementation Details

The whole network is trained on a Nvidia GTX 1060 GPU and Intel Core i5-8400 2.80GHz CPU with Tensorflow[13]. We train our network on the LOL dataset and use the SGD technique to optimize our results. For the Decom-net, we set the batch size 10 and patch-size 48x48. While for the Restore-Net and Adjust-Net, we set the batch size 4 and patch-size 384x384. In experiments, we pretrain the reinforce network using LOL dataset as input and fine-tune the network using images generated by the enhancement module. For the reinforce-net, we set the batch size 32 and patch-size 128x128.

3.2. Quantitative Evaluation

As shown in Table 1, we compare our approach with different methods on LOL dataset. Besides, we also evaluate our method on LIME [8], NPE [5], and MEF [19] datasets as shown in Figure 4. We also adopt three metrics to evaluate image quality, which are PSNR, SSIM and NIQE [20]. As for PSNR
and SSIM, the value is higher that means the image quality is better. However, as for NIQE, the lower value means the image quality is better.

Figure 4. The results of NIQE on different datasets. From the results, we see that our MSIGEN is the lowest among all the other methods in LIME and NPE.

Table 1. The results of PSNR, SSIM, NIQE on LOL dataset. From the results, we see that our MSIGEN is highest among all the other methods in PSNR and SSIM. In terms of NIQE, our MSIGEN also is lower than most of the other methods. It suggests that exchanging partial information and combining illumination information enable our MSIGEN model to make good use of global features from the “normal”-light images and local features from the illumination information.

4. Conclusions
In this work, we propose a novel two-stage enhanced network to achieve low-light image enhancement, named MSIGEN. The enhancement module recovers well-exposed image details and decreases noise variance and color bias by regression. Also, we introduce a reinforce module that refine the initial images to effectively remove visual defects amplified and improve the details. Our experimental results have shown that our model performs better than the other approaches in PSNR and SSIM. The proposed method generates high-quality images with abundant textures and naturalness. In the next step, we plan to use our model to real-time processing, and design a more powerful network in real-time processing and apply the model to other enhancement tasks (e.g., low-light video enhancement).

Acknowledgments
This project is supported by National Natural Science Foundation of China (Grant No. 52075404) and Application foundation frontier special project of Wuhan Science and Technology Bureau (No.2020010601012176).

Reference
[1] Pizer S M, Amburn E P, Austin J D, et al. Adaptive histogram equalization and its variations[J]. Computer Vision Graphics & Image Processing, 1987, 39(3):355-368.
[2] Land, Edwin. The retinex theory of color vision[J]. Scientific American, 1977, 237(6):108.
[3] Jobson D J, Rahman Z. Properties and performance of a center/surround retinex[J]. IEEE Transactions on Image Processing, 1997, 6(3):P.451-462.
[4] Jobson, D. J, Rahman, et al. A multiscale retinex for bridging the gap between color images and...
the human observation of scenes[J]. IEEE Transactions on Image Processing, 2002, 6(7):965–76.

[5] Wang S, Zheng J, Hu H M, et al. Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2013, 22(9).

[6] Fu X, Zeng D, Huang Y, et al. (2016) A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp .2782–2790.

[7] Lore K G, Akintayo A, Sarkar S. LLNet: A Deep Autoencoder Approach to Natural Low-light Image Enhancement[J]. Pattern Recognition, 2017, 61:650-662.

[8] C. Wei, W. Wang, W. Yang, and J. Liu, Deep retina decomposition for low-light enhancement [J], in British Macine Vision Conference, 2018.

[9] Dabov K, Foi A, Katkovnik V, et al. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering[J]. IEEE Transactions on Image Processing, 2007, 16(8):2080-2095..

[10] Zhang Y, Zhang J, Guo X. Kindling the Darkness: A Practical Low-light Image Enhancer[C]// 2019.

[11] O. Ronneberger, P. Fischer, and T. Brox, (2015)U-net: Convolutional networks for biomedical image segmentation, in MICCAI, pp. 234–241,

[12] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In ICML, 2010. 4

[13] Xu J, Zhang L, Zhang D, et al. Multi-channel Weighted Nuclear Norm Minimization for Real Color Image Denoising[C]// 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017.

[14] Z. Ying, L. Ge, and W. Gao, A bio-inspired multi-exposure fusion framework for low-light image enhancement, arXiv, 2017.

[15] Z. Ying, L. Ge, Y. Ren, R. Wang, and W. Wang, (2017) A new low-light image enhancement algorithm using camera response model, in IEEE International Conference on Computer Vision Workshop,

[16] X. Dong, Y. Pang, and J. Wen, (2011)Fast efficient algorithm for enhancement of low lighting video, in IEEE ICME, pp. 1–6.

[17] X. Fu, D. Zeng, H. Yue, et al. A fusion-based enhancing method for weakly illuminated images[J], Signal Processing, 2016, vol. 129, pp. 82–96

[18] W. Wang, W. Chen, W. Yang, and J. Liu,(2018). GLADNet: Low-Light Enhancement Network with Global Awareness. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE Computer Society.

[19] L. Chulwoo, L. Chul, L. Young-Yoon, and K. Chang-Su, Power-constrained contrast enhancement for emissive displays based on histogram equalization[J], IEEE Trans Image Process, 2012, vol. 21, no. 1, pp. 80–93

[20] A. Mittal, R. Soundararajan, and A. Bovik. Making a completely blind image quality analyzer[J], IEEE Signal Processing Letters,2013, vol. 20, no. 3

[21] Guo X, Li Y, Ling H. LIME: Low-light Image Enhancement via Illumination Map Estimation[J], IEEE Transactions on Image Processing, 2016, PP(99):1-1