

# UCLA CS MS Project: Low-light Illumination and Reflectance based Image Enhancement

Yinxue (Yolanda) Xiao  
University of California, Los Angeles  
yolandaxiao7@gmail.com

May 28, 2020



Figure 1: Examples of the Low-Light Image Enhancement Results using LLIR-Net

## Abstract

Capturing well illuminated photographs in low-light conditions has been a challenging task. Low-light environments have low signal-to-noise ratio and result in photos with poor visual quality. In order to enhance the low-light images, we propose an end-to-end network based on Retinex theory. Retinex theory assumes that an image can be decomposed to the product of two images: illumination and reflectance. Our system consists of two networks that enhance the two images respectively, and combined them to produce the final enhanced image. Compared to other state-of-the-art low-light image enhancement methods, our approach produces better-illuminated results both qualitatively and quantitatively.

## 1. Introduction

Photos are an essential part of people's daily lives - they are used to capture memorable moments, to document great scenery, and to digitize important documents. However, it's often hard to obtain the ideal photo when there's not enough

lighting, for example, when it's at night or when an object is shadowed. Thus, it would be beneficial for users to have a software that automatically enhances images with low-light.

Producing high quality photographs in low light environments has been a challenging task. Due to insufficient lighting, captured images suffer from low signal-to-noise ratio. Traditional approaches for low-light image enhancement include histogram-based methods [7, 18, 17], Retinex-based methods [9, 6] and filtering-based methods. More recently, researchers have focused mainly on learning-based methods such as Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs).

GANs [8] include both a generator network that aims to create enhanced images and a descriptor network that aims to differentiate the enhanced and targeted images. Its design of adversarial loss makes it powerful in generating new images almost indistinguishable from the target ones. However, although two-way GANs have less requirement on the input dataset, it is computationally expensive and hard to train, as it requires training 4 different networks.

CNN approaches [4, 2, 25] use various encoder-decoder based networks to obtain the enhanced images. Works such

as [4, 25] incorporate Retinex theory to the convolutional neural networks design for improved results.

Inspired by Retinex theory, we propose a **Low-Light Illumination and Reflectance Network (LLIR-Net)**. It consists of 1) a module that decomposes an image to illumination and reflectance based on morphological closing, and 2) two networks that enhance the illumination and reflectance of the input images respectively. On top of the traditional  $L_1$  and  $L_2$  losses, we incorporated perceptual loss based on image content information, and total variation loss that smooths out the illumination.

## 2. Related Works

**Traditional Methods based on Retinex Theory** In Retinex theory [12], an image can be decomposed into the product of reflectance and illumination. LIME[9] estimates the illumination map by first finding the maximum value of each pixel in all channels and refine it by imposing a structure prior. Fu et al. proposed a fusion-based method [6] for enhancing weakly illuminated images by fusing multi-scale inputs and weights to obtain the adjusted illumination. These methods only estimate the illumination and compute the reflectance by dividing the original image with the illumination. While they use different techniques to improve the illumination, the reflectance is assumed to be accurate to be combined with the adjusted illumination for the enhanced result. However, the computed reflectance is not noise-free. The reflectance decomposed from low-light images exhibit much more noise than that of the normal-light images. The noise from the reflectance of the low-light images will thus results in noise in the enhanced images.

**Generative Adversarial Networks (GANs)** CycleGAN[26] stacks two GANs together along with a cycle consistency loss to achieve training on unpaired image data. It applies to a wide variety of applications in image-to-image translation. Chen et al. use a 2-way GAN based method [3] with improved global U-Net and Wasserstein GAN (WGAN) to produce enhanced photographs. Jiang et al. proposed EnlightenGAN [10], which is a 1-way GAN without paired supervision with a global-local discriminator structure, perceptual loss, and attention incorporated. While these methods allow unpaired training, they are computational expensive and hard to train as 2 or 4 networks are involved.

**Convolutional Neural Networks (CNNs)** Chen et al. in [2] use an end-to-end network with incorporated noise and global/local information to obtain enhanced images from extremely low-light RAW images. However, the RAW data format itself has more advantages in image enhancement as it contains more information compared to that of the JPGs or PNGs. Several approaches incorporate the traditional Retinex theory based ideas with CNNs to obtain enhanced images from end-to-end training. RetinexNet [4] use a de-

composition network to decompose input images into reflectance and illumination, and then an encode-decoder network to adjust the illumination. Zhang et al. in [25] extends RetinexNet by adding another reflectance restoration network to improve on the reflectance. It includes an image decomposition network, a illumination adjustment network, as well as a reflectance restoration network. The adjusted illumination and improved reflectance are combined to produce the final enhanced image.

## 3. Method

The proposed approach includes three parts: 1) an image decomposition step, 2) an illumination adjustment network, and 3) a reflectance adjustment network. The method overview is shown in Fig.2. The image decomposition step decomposes a low-light input image into reflectance and illumination. The illumination adjustment network enhance the illumination, and reflectance adjustment network denoise the reflectance. The final illumination and reflectance images are combined to produce the enhanced lighting image.

### 3.1. Illumination Estimation

In Retinex theory, an image can be expressed as the the product of illumination and reflectance. The equation is

$$S^c(x, y) = R^c(x, y) \cdot I(x, y) \quad (1)$$

where  $S$  is the input image,  $R$  is the reflectance,  $I$  is the illumination, and  $c$  is the color channel. The pixel position is denoted by  $(x, y)$ . Based on this equation, the problem of enhancing a low light image directly transforms into computing the illumination and reflectance of the enhanced image. In ideal situation, the reflectance of low-light and normal-light should be the same, while the illumination gives the lighting difference. The goal thus become estimating the illumination of enhanced images from that of the low-light input images.

A naive computation of the initial illumination is to find the maximum pixel value from the three channels. The corresponding reflectance can be obtained by reversing the Eq.1 [9]. Note that  $\epsilon$  is added to avoid dividing by zero.

$$L(x, y) = \max_{c \in \{R, G, B\}} S^c(x, y) \quad (2)$$

$$R(x, y) = S(x, y) / (L(x, y) + \epsilon) \quad (3)$$

An improved initial illumination estimation method [6] is to use the morphologically closing operation to enhance the illumination. The operation smooths images by closing small holes and black points in images, thus creating more realistic illumination results.

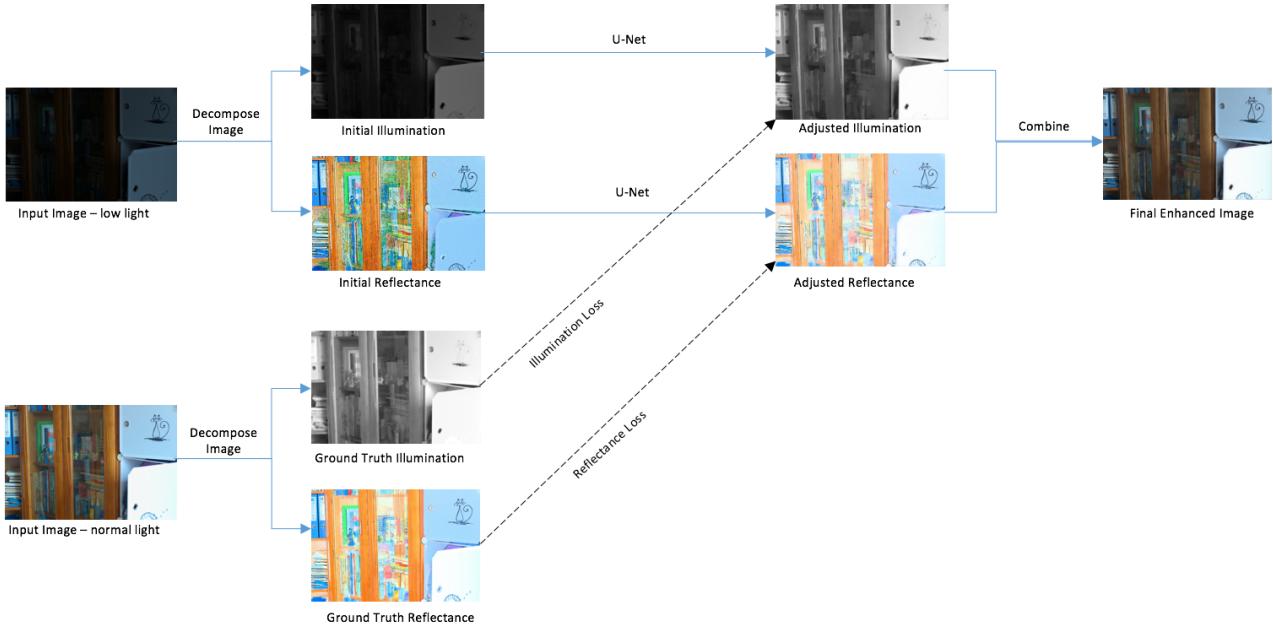


Figure 2: Overview of the Low-light Image Enhancement Pipeline

In this project, the initial illumination estimation uses 1) the maximum pixel value of all channels, and 2) the morphologically closing operation. The reflectance estimation is calculated accordingly based on Eq.3.

### 3.2. Illumination Adjustment Network

Given the initial illumination of the low-light images, we aim to find the adjusted illumination for normal light conditions. The network used is based on U-Net, with 7 times of downsampling and upsampling, and 4x4 convolutions. We incorporated total variation loss to smooth out the illumination. The total weighted loss function is shown in Eq. 4.

$$\begin{aligned} \ell_{illumination}(\hat{I}, I) = & \lambda_1 \cdot \ell_1 + \lambda_2 \cdot \ell_2 \\ & + \lambda_{TV} \cdot \ell_{TV}(\hat{I}) \\ & + \lambda_{MS-SSIM} \cdot \ell_{MS-SSIM}(\hat{I}, I) \end{aligned} \quad (4)$$

### 3.3. Reflectance Adjustment Network

While most traditional illumination-based method assume the reflectance is the same among images of different lighting conditions, the reflectance computed from the Retinex theory for low-light and normal light images are in fact different. An example can be seen in Fig.3. The reflectance of the low-light image includes a lot of noise. The reflectance adjustment network is thus designed to denoise the reflectance and improve the color of the low-light image.

The network is also based on U-Net, with 7 downsampling and upsampling steps, and 4x4 convolutions. We in-

corporated perceptual loss to match the content of the enhanced and target images. The total weighted loss function is shown in Eq. 5.

$$\begin{aligned} \ell_{reflectance}(\hat{I}, I) = & \lambda_1 \cdot \ell_1 + \lambda_2 \cdot \ell_2 \\ & + \lambda_{perceptual} \cdot \ell_{perceptual}(\hat{I}, I) \\ & + \lambda_{MS-SSIM} \cdot \ell_{MS-SSIM}(\hat{I}, I) \end{aligned} \quad (5)$$

### 3.4. Network Architecture

Both networks use the same architecture that is based on U-Net with skip connection blocks [19] and follows the design used in [27]. Different from the original U-Net design with convolutions and max pooling, strided convolutions are used for downsampling and upsampling. The images are trained with batch size = 1, and patches of 384 x 384. The input and output for the networks are of size [channel] x 384 x 384, with illumination having 1 channel, and reflectance having 3 channels. The inputs are downsampled 7 times into 512 x 3 x 3, and then upsampled 7 times with concatenated parts to obtain the output size. The downsampling convolution uses a kernel of size 4 x 4 with stride 2. The network architecture is shown in Fig.4. Note that the first and last convolution step is not followed by batch normalization, and the last convolution step uses ReLU instead of leaky ReLU. The last deconvolution is followed by Tanh activation to produce the final result.

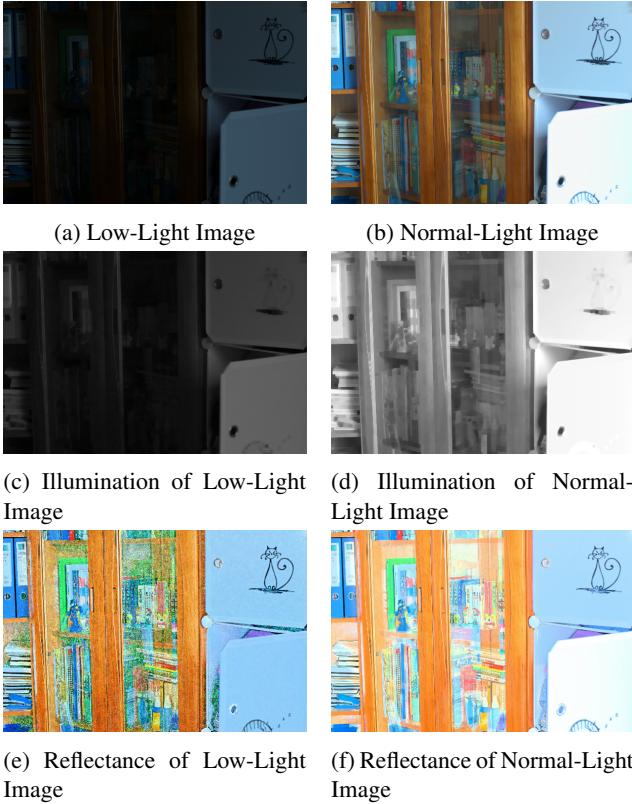


Figure 3: Illumination and Reflectance of Low-Light and Normal-Light Image

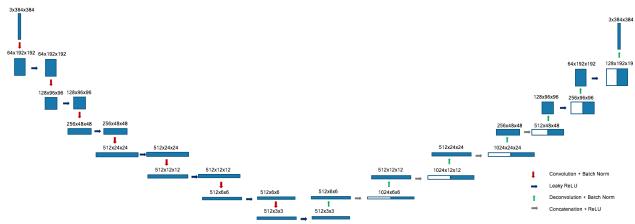


Figure 4: Architecture of the U-Net we used

### 3.5. Loss Functions

The networks use several different loss functions including: 11 loss, l2 loss, total variation loss, perceptual loss [11], and multi-scale structural similarity index (MS-SSIM) loss [23].

**Perceptual Loss** Perceptual loss measures the Euclidean distance between the feature representations of two images, which is essentially the feature reconstruction loss. Comparing the feature maps of two images encourages the content of the output image to match that of the target image's, instead of matching the exact pixel values. In this project, a 16-layer VGG network [21] pretrained on the ImageNet dataset [20] is used to compute an image's feature representation.

The equation for computing the perceptual loss is shown in Eq.6.

$$\ell_{perceptual}(\hat{I}, I) = \|\phi_j(\hat{I}) - \phi_j(I)\|_2^2 \quad (6)$$

$I$  denotes the target image, while  $\hat{I}$  denotes the network output image.  $\phi_j$  is the  $j$ th layer of the network  $\phi$ .

**Total Variation Loss** The total variation loss regularizes the x and y gradients of the image, thus encouraging spatial smoothness in the output images. The equation is shown below in Eq.7.

$$\ell_{TV}(\hat{I}) = \|\nabla_x \hat{I}\|_2^2 + \|\nabla_y \hat{I}\|_2^2 \quad (7)$$

### Multi-Scale Structural Similarity (MS-SSIM) Loss

[23] SSIM is a reference-based metric that measures the similarity between two images, focusing on local structures of the image. MS-SSIM extends it by combining weighted SSIM at different scales. The equation for computing the MS-SSIM loss is shown in Eq.8.

$$\ell_{MS-SSIM}(\hat{I}, I) = 1 - MS-SSIM(\hat{I}, I), \quad (8)$$

## 4. Experiments

### 4.1. Dataset

There are several datasets specifically for the low-light image enhancement task. The Exclusively Dark (ExDark) Dataset [14] contains 7,363 low-light images from 10 different low-light environments conditions with 12 object classes. However, it doesn't contain corresponding normal-light images, which is required to paired image training. The MIT-Adobe FiveK Dataset [1] includes 5,000 photographs taken with SLR cameras in RAW format. It also includes corresponding edited images from five photography professionals. Even though the task is similar to low-light image enhancement, some image pairs focus on style instead of lighting improvements. The See-in-the-Dark (SID) Dataset, collected by [2], contains 5094 RAW short-exposure, long-exposure image pairs and includes both outdoor and indoor images with SLR cameras. However, the intensities of the low-light images captured in this dataset are less than 5 lux. It is thus hard to restore the low-light images if they are transformed to jpg images since it contains too little information in extremely dark situations.

This project uses the LOw-Light (LOL) dataset [4] and a synthetic dataset prepared by [4] as well. The LOL dataset contains 500 low/normal-light image pairs from a variety of scenes. The paired images are captured using the same camera with different exposure time and ISO. The synthetic dataset contains 1000 images from RAISE [5], and their corresponding synthesized low-light images. The histograms of the Y channel in YCbCr space of the synthesized low-light images match that of the captured low-light images in real life.

Three datasets are used for testing: the DICM[13], MEF[15] and NPE[22] dataset. The DICM dataset contains 69 images from commercial digital cameras, and we select the first 44 for the low-light image enhancement task. The MEF dataset includes 17 images with different exposure levels. The NPE dataset contains 8 underexposed images of the outdoor scenes.

## 4.2. Experimental Setup

The proposed method is implemented in PyTorch 1.5 with Python 3.6. The training is run on Ubuntu 16.04 as well as Google Colaboratory. The Ubuntu machine has Intel Core i9-7920X 2.90GHz CPU and a NVIDIA Titan V GPU. Google Colaboratory has Intel Xeon 2.00GHz CPU and a Tesla P4 GPU. Testing is performed on macOS with Intel Core i5 2.9 GHz CPU. Matlab r2020a is used to compute the PSNR, SSIM and NIQE results.

## 4.3. Training Details

The network is trained on the LOL dataset [4] and the synthetic dataset prepared by [4]. The 500 images in the LOL dataset are divided into a training set with 480 images and an evaluation set with 15 images. A total of 1480 image pairs are trained on the proposed network for 200 epochs. Learning rate decay is applied to the last 100 epochs. Adam is used as the optimizer with a learning rate of 0.0002. For the perceptual loss, it is computed from the `relu2_2` layer of the VGG-16 network. The hyper-parameters in the loss function 4 for the illumination network are set to  $\lambda_1 = 100$ ,  $\lambda_2 = 100$ ,  $\lambda_{MS-SSIM} = 100$ ,  $\lambda_{TV} = 0.0001$ . The hyper-parameters in the loss function 5 for the reflectance network are set to  $\lambda_1 = 100$ ,  $\lambda_2 = 100$ ,  $\lambda_{perceptual} = 0.0001$ ,  $\lambda_{MS-SSIM} = 100$ .

## 4.4. Evaluation Metrics

This project uses three different evaluation metrics to evaluate the performance of the network: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) [24], and Natural Image Quality Evaluator (NIQE) [16].

Objective image quality assessment (IQA) evaluates the distortions and degradations of an image. It includes full reference (FR) methods, where a reference image is supplied to help assess the test image's quality, and no reference (NR) methods, where no reference image is used. For NR IQA methods, it can be divided into opinion-aware (OA) and opinion unaware (OU) methods, depending on if the training dataset is with subjective input from human. Within the OU NR IQA methods, it can be further divided into distortion-aware (DA) and distortion-unaware (DU), depending on if the image distortions are known or not. The NIQE is a OU-DU NR IQA method, while PSNR and SSIM are FR IQA approaches. Both FR and NR based IQA methods are used for more comprehensive comparison.

**Peak Signal-to-Noise Ratio (PSNR)** PSNR computes the peak signal-to-noise ratio between two images. It first calculates the mean-square error (MSE) between the test and target images, and use the ratio of the squared maximum pixel value divided by MSE. It is in logarithmic decibel scale since there's a wide dynamic range for many signals. The equation is shown below.

$$MSE = \frac{\sum_{x=1}^W \sum_{y=1}^H (\hat{I}[x, y] - I[x, y])^2}{W \cdot H} \quad (9)$$

$$PSNR = 10 \log_{10} \frac{R^2}{MSE} \quad (10)$$

R denotes the maximum pixel value of the target image  $I$ .  $[x, y]$  denotes the pixel location.

**Structural Similarity (SSIM) Index** [24] SSIM is a image quality assessment method based on the degradation of structural information. It is based on the assumption that the human visual system (HVS) is more sensitive in extracting structural information. It computes the product of the luminance comparison  $l(x, y)$ , the contrast comparison  $c(x, y)$  and structure comparison  $s(x, y)$  functions.

$$\begin{aligned} SSIM(x, y) &= l(x, y) \cdot c(x, y) \cdot s(x, y) \\ &= \frac{(2\mu_x\mu_y + c_1)}{(\mu_x^2 + \mu_y^2 + c_1)} \cdot \frac{(2\sigma_x\sigma_y + c_2)}{(\sigma_x^2 + \sigma_y^2 + c_2)} \\ &\cdot \frac{(\sigma_{xy} + c_3)}{(\sigma_x\sigma_y + c_3)} \\ &= \frac{(2\mu_x\mu_y + c_1) \cdot (2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1) \cdot (\sigma_x^2 + \sigma_y^2 + c_2)} \end{aligned} \quad (11)$$

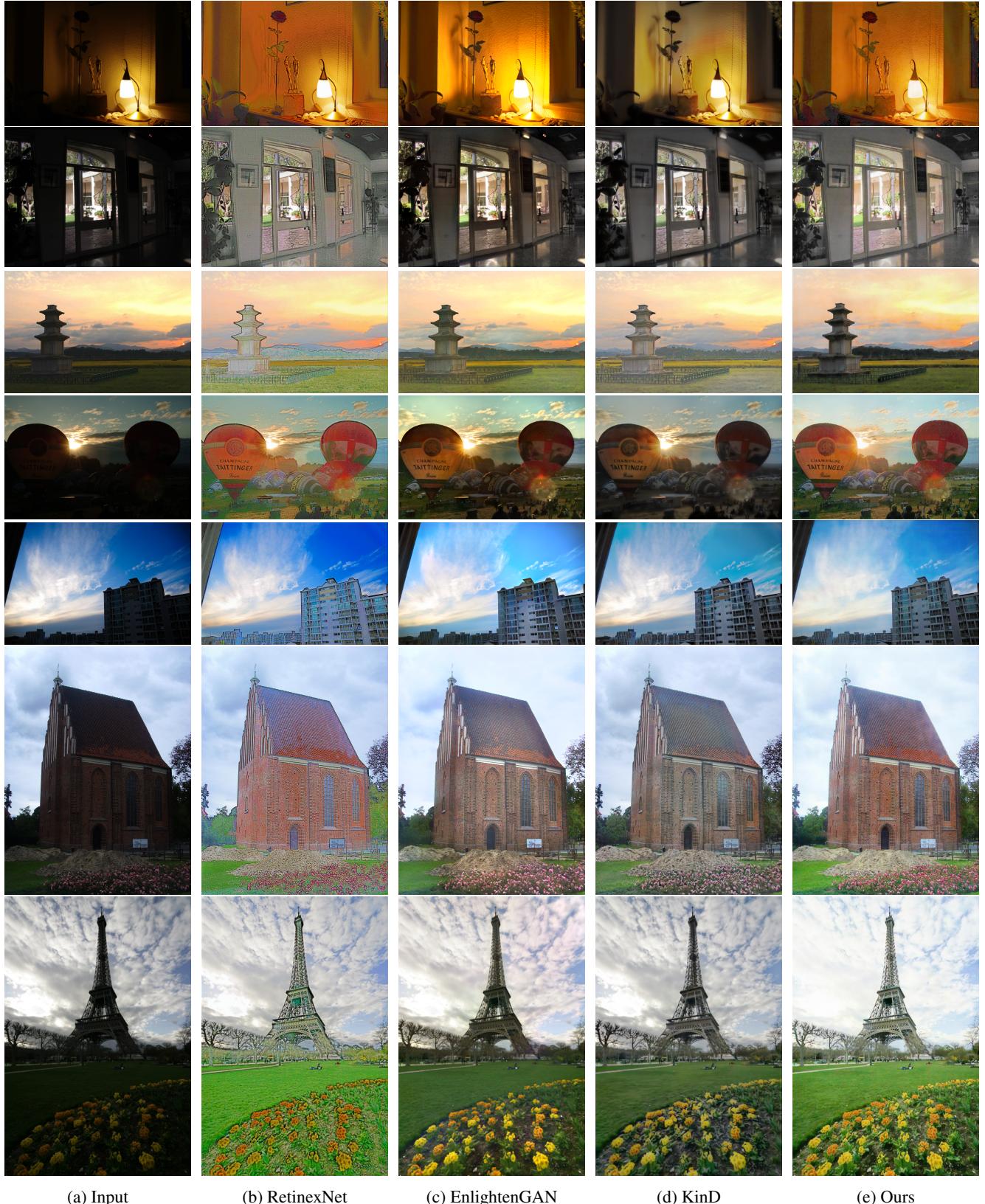
In Eq. 11,  $x$  and  $y$  denote two images.  $\mu_x$  and  $\mu_y$  are the mean intensities for the two images,  $\sigma_x$  and  $\sigma_y$  are the standard deviations of the images. The luminance comparison  $l(x, y)$  is the comparison of the mean intensities, the contrast comparison  $c(x, y)$  is the comparison of the standard deviations, and the structure comparison  $s(x, y)$  is the comparison of the normalized images. Note that the equation can be simplified from step 2 to step 3 by setting  $c_3 = c_2/2$ .

**Natural Image Quality Evaluator (NIQE)** NIQE is a 'completely blind' image quality analyzer that is based on statistical regularities in natural images. It uses a space domain natural scene statistic (NSS) model to compute a collection of statistical features of natural images, and fits them to a multivariate Gaussian (MVG) model. The quality of the test image is computed as the distance between the features of the test image and that of a collection of natural images.

## 5. Results and Discussion

### 5.1. Qualitative Results

Figure. 5 shows the visual comparisons on state-of-the-art low-light image enhancement methods on the DICM[13]



(a) Input

(b) RetinexNet

(c) EnlightenGAN

(d) KinD

(e) Ours

Figure 5: Visual Comparison on the Low-light Image Enhancement Methods

and MEF[15] datasets. The left most column shows the low-light input image. The second to fourth columns show enhanced images by Retinex-Net [4], EnlightenGAN [10], and KinD [25] respectively. The last column shows results from our proposed network.

Retinex-Net tends to produce cartoon like results: the objects in the images have distinct boundaries and the color is overly vibrant. For example, the house on the 6th row and the Eiffel Tower on the 7th row looks over saturated. CycleGAN and KinD produce more reasonably enhanced images that's similar to ours.

KinD is good at denoising the dark regions in images. However, it tends to over smooth it. In the blue sky building picture on the 5th row, for example, the windows on the building in the bottom right corner are very blurry. In contrast, the building in our picture is more clear, although slightly more noisy. The same issue can be seen in the flower in the 1st row, and the wall/windows in the 2nd row.

EnlightenGAN seems to perform better than Retinex-Net and KinD. However, our approach yields more reasonable and appealing enhancements. For example, in the indoor images in 1st and 2nd row, our method enhance the darker regions more and brings the whole image to a appropriate illumination level. It also balance out the dark and bright regions. For the outdoor images in the 4th, 6th and 7th row, our methods gives a slightly brighter and more vibrant color, which is more aesthetically pleasing.

## 5.2. Quantitative Results

Table. 1 reports the training PSNR, SSIM and NIQE results on U-Net, Retine-Net [4], EnlightenGAN [10], KinD [25], and our netowrk on the LOL evaluation dataset.

Metrics	EnlightenGAN	Retinex-Net	KinD	LLIR-Net(Ours)
PSNR	17.526	16.774	20.379	<b>22.566</b>
SSIM	0.676	0.566	0.807	<b>0.823</b>
NIQE	4.880	9.729	3.985	<b>2.819</b>

Table 1: Training Results on LOL dataset based on PSNR, SSIM, and NIQE metrics

Methods	DICM	MEF	NPE
Retinex-Net	4.712	4.905	4.066
EnlightenGAN	2.767	2.913	3.317
KinD	2.991	3.375	3.529
LLIR-Net(Ours)	<b>2.668</b>	<b>2.861</b>	<b>2.939</b>

Table 2: Quantitative Results on DICM, MEF, NPE datasets based on the NIQE metric

Table. 2 reports the NIQE results on Retinex-Net [4], EnlightenGAN [10], KinD [25], and our network on the DICM[13], MEF[15] and NPE[22] datasets. The NIQE metrics is used on these datasets because there are no suit-

able paired low-light normal-light image datasets for testing on FR IQA standards. It shows that our method yields the lowest NIQE value, which means it produces the most realistic image enhancement results out of all the methods evaluated below.

## 6. Conclusion

In this project, we have proposed an end-to-end low-light illumination and reflectance based image enhancement network, called LLIR-Net. It's based on the idea of Retinex theory, where an image can be decomposed to illumination and reflectance. Our approach consists of two networks that enhance the illumination and reflectance of an input image respectively. The experiments show that our proposed network outperforms multiple state-of-the-art methods both qualitatively and quantitatively.

In future work we hope to explore ways of more effective denoising in darker images, as well as producing enhanced images with adjustable illumination levels.

## References

- [1] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input / output image pairs. In *The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition*, 2011.
- [2] 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.
- [3] Y. Chen, Y. Wang, M. Kao, and Y. Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6306–6314, 2018.
- [4] Wenhan Yang Jiaying Liu Chen Wei, Wenjing Wang. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*, 2018.
- [5] Duc-Tien Dang-Nguyen, Cecilia Pasquini, Valentina Conotter, and Giulia Boato. Raise: a raw images dataset for digital image forensics. In *Proceedings of the 6th ACM Multimedia Systems Conference*, pages 219–224, 2015.
- [6] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xing-hao Ding, and John Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Process.*, 129(C):82–96, December 2016.
- [7] Rafael C Gonzales and Richard E Woods. *Digital image processing*, 2002.
- [8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q.

- Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.
- [9] X. Guo, Y. Li, and H. Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017.
- [10] 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.
- [11] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.
- [12] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.
- [13] 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.
- [14] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42, 2019.
- [15] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015.
- [16] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2012.
- [17] 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.
- [18] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderweld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.
- [19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [20] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.
- [21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [22] 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.
- [23] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, 2003.
- [24] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [25] 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.
- [26] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.
- [27] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.