

Seeing in the Dark by Component-GAN

Ning Rao, Tao Lu , *Member, IEEE*, Qiang Zhou, Yanduo Zhang , *Member, IEEE*,
and Zhongyuan Wang , *Member, IEEE*

Abstract—Recently, Retinex theory based low-light image enhancement (LLIE) algorithms have achieved impressive results in controlled environment. However, the majority of deep learning based LLIE algorithms leverage relighting by enhancing the illumination components that directly determines the image brightness, regrettably, they ignore the information of reflectance components, which may cause problems such as image noise and color distortion in reconstructed images. To tackle this problem, in this letter, we propose a component enhancement network based on Generative Adversarial Network (Component-GAN) for recovering clear images from low-light ones. Specifically, the network is composed of the decomposition part for dividing the paired low/normal-light images into illumination components and reflectance components, and the enhancement part for generating high-quality images. It is worth to note that we provide two branches of component enhancement network, which are parallel to improve the two components simultaneously. Hereby, we treat the reconstruction part as the generative network and adopt discriminative network to boost image reconstruction performance. Through extensive experiments, the proposed approach outperforms some state-of-the-art LLIE methods in terms of visual and subjective qualities.

Index Terms—Component enhancement network, generative adversarial network, light enhancement, low-light image.

I. INTRODUCTION

IMAGE captured in the dark suffers from poor visibility, low contrast and high ISO noise. Low-light images not only affect people's ability to perform visual tasks, but also bring great challenges in various machine-learning tasks such as face recognition, pedestrian detection and automatic driving [1]. Therefore, it is necessary to improve the quality of low-light images. In the past few decades, there have been a lot low-light images enhancement (LLIE) algorithms devoted to solve this problem. Generally, it can be divided into two categories: mathematical model based and deep learning based methods.

Manuscript received March 5, 2021; revised April 28, 2021; accepted May 2, 2021. Date of publication May 12, 2021; date of current version June 28, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 62072350, U1903214, 62001334, 61771353, and 62071339; in part by the Hubei Technology Innovation Project under Grant 2019AAA045; in part by the Central Government Guides Local Science and Technology Development Special Projects under Grant 2018ZYYD059; in part by the High value Intellectual Property Cultivation Project of Hubei Province; in part by the Enterprise Technology Innovation Project of Wuhan under Grant 202001602011971. (Ning Rao and Tao Lu contributed equally to this work and are co-first authors.) (Corresponding author: Tao Lu.)

Ning Rao, Tao Lu, Qiang Zhou, and Yanduo Zhang are with the Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430200, China (e-mail: 1353935562@qq.com; lutxy1@gmail.com; 315653752@qq.com; zhangyanduo@hotmail.com).

Zhongyuan Wang is with the School of Computer Science, Wuhan University, Wuhan 430072, China (e-mail: wzy_hope@163.com).

Digital Object Identifier 10.1109/LSP.2021.3079848

Mathematical model based methods use statistical model to adjust the light intensity distribution for better visual results. There are two categories of the most representative methods: one is based on Retinex theory [2], [3] and another is based on histogram equalization and its variants [4]–[6]. Based on the logarithmic transformation, Fu *et al.* [7] proposed a weighted variational model to estimate both the reflectance and illumination component from the observed image with imposed regularization terms. Guo *et al.* [8] enhanced the illumination component by imposing priors. Fu *et al.* [9] proposed a fusion-based method to enhance the illumination component by fusing original illuminance with luminance-improved and contrast-enhanced versions. Park *et al.* [10] presented an algorithm that enhances undesirably illuminated images by generating and fusing multi-level illuminations from a single image. The final image is obtained by multiplying the equalized illumination and enhanced reflectance. However, due to the nonlinearity across color channels and data complexity, existing methods have limited capabilities to enhance color, since color is easy to be distorted locally.

Deep learning based methods developed Convolutional Neural Networks (CNNs) to view low-light images as image reconstruction problems. Gharbi *et al.* [11] proposed HDR-Net, which incorporated deep networks with the ideas of bilateral grid processing and local affine color transforms with pairwise supervision. RetinexNet [12] combined Retinex theory with CNNs to estimate illuminance map and enhance low illumination images by enhancing illuminance map. Li *et al.* [13] proposed a new Retinex model that takes noise into consideration. Inspired by GAN, Jiang *et al.* [14] used unsupervised manner to learn a GAN to yield good results. Chunle *et al.* [15] proposed a zero-reference deep curve estimation to recover low-light images.

As we known, the illumination component represents the various lightness on objects. While for low-light images, it usually suffers from darkness and unbalanced illumination distributions. So most LLIE approaches heavily rely on enhancing the illuminance channel. In order to suppress the low-light image noise, some methods just cascade denoising and lighting enhancement modules to achieve better performance. Neither denoising before enhancement (may result in blurring) nor enhancement before denoising (may cause noise amplification) own perfect relighting performance.

Taking into account the above issues, we discover that reflection component always benefits texture reconstruction. Because reflection component describes the intrinsic property of the observed objects, which is considered to be consistent under any lightness conditions. Details of the low-light images are retained on the reflectance and some texture parts of images may degrade by noise from reflectance channel. So enhancing reflectance also matters.

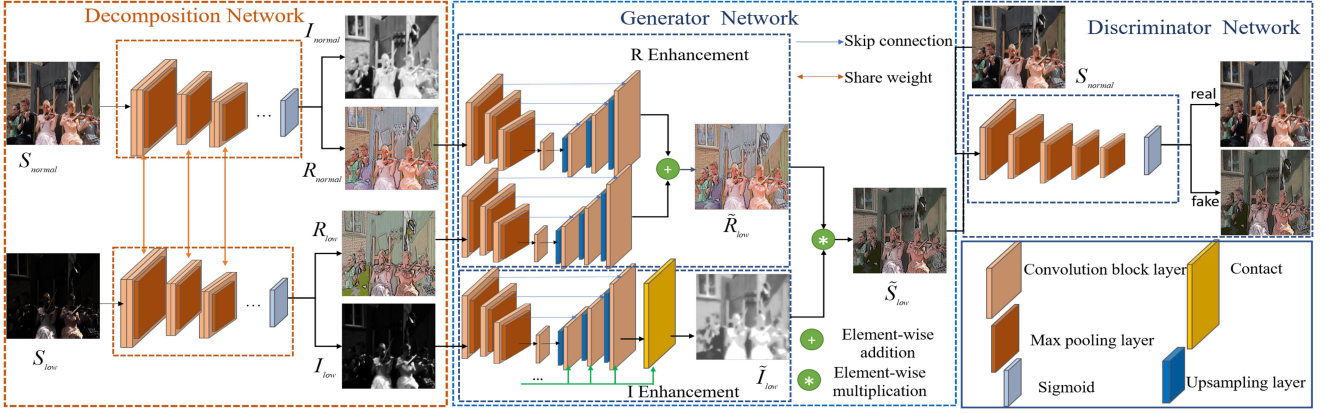


Fig. 1. The framework of Component-GAN. Component-GAN contains decomposition network, generator network and discriminator network. Decomposition network decomposes the low/normal-light images into reflectance and illumination channels. Generator network enhances two components respectively, and then fuse these two channels together for rendering an enhanced output. Here, U-Net is adopted as the backbone. Then we use the discriminator of VGG-19 architecture to distinguish the generated image according to the normal light image to obtain better image reconstruction quality.

Based on this observation, in this letter, we propose a component enhancement network based on Generative Adversarial Network (Component-GAN) to cope with these problems. First, we input the paired low/normal-light images into the decomposition network, in which the images will be divided into reflectance components and illumination components. Second, we design two branch enhancement networks for the retinex components, and then fuse them together to reconstruct a light enhancement image. Finally, discriminator is adopted to judge whether the rendered images meet the reconstruction criterion. To summarize the major contributions of this work: 1) We propose two branch component enhancement networks to optimize both illumination and reflectance simultaneously. We further develop the power of reflectance. The combination of these two components successfully reduces most noise and achieves impressive performance. 2) In order to make full use of paired low/normal-light images, we propose a component-GAN method to extend the Retinex network into GAN scheme. The designed discriminator further reduces noise and artifacts. 3) We perform evaluations on the proposed method compared with some popular datasets, which demonstrates the superiority of the proposed method qualitatively and quantitatively.

II. THE PROPOSED METHOD

A. Network Architecture

As shown in Fig. 1, the network architecture of component-GAN has three parts, which will be described as follows:

1) *Decomposition Network*: The network consists of two sub-networks with same structure and same parameters. Each sub-network has only 5 convolutional layers. S_{normal} and S_{low} represent the normal-light input image and the low-light input image. Through the sub-networks, they will be decomposed into the reflectance and illumination components, which means we will get R_{normal} , I_{normal} from S_{normal} and R_{low} , I_{low} from S_{low} .

2) *Generator Network*: It is composed of reflectance and illumination enhancement module. Different from RetinexNet [12], we add U-Net in illumination enhancement module to reduce noise and artifacts. Considering U-Net [16]

has achieved huge success on image enhancement, we adopt it as our enhancement network backbone. We also find that the reflection component R_{normal} has the effect of constraining its paired low-light reflection component R_{low} . Actually, both components as reflection and illumination channels from normal light images can boost the low-light components reconstruction performance. Finally, we fuse the components of two branches to generate a reconstructed image, which is the input of discriminator.

3) *Discriminator Network*: We use VGG19 network with FC layers to distinguish the reconstructed image \tilde{S}_{low} in line with the real normal light image S_{normal} . The discriminator further improves the reconstruction quality.

B. Loss Function

In order to learn the mapping function of low/normal-light images through N image pairs $\{S_{low_i}, S_{normal_i}\}_{i=1}^N$, in which i is the index of sample pairs, we design a loss function L . The loss L consists of four components: illumination components loss L_I , invariable reflectance components loss L_R , reconstruction loss L_{Rec} and GAN loss L_G . It is represented as:

$$L = \lambda_I L_I + \lambda_R L_R + \lambda_{Rec} L_{Rec} + \lambda_G L_{GAN}, \quad (1)$$

where λ_I , λ_R , λ_{Rec} , λ_G are the balance parameters for controlling the contributions of each loss term.

A good decomposition method should consider the decomposed illumination components preserving the detailed texture and overall structure. In order to make that comes true, we adopt an improved TV loss [12] as the illumination component loss. Total variation minimization [17] minimizes the gradient of the whole image. It is often used to preserve the overall structure and texture information. The L_I is formulates as:

$$L_I = \sum_{i=1}^N \|\nabla I_{low_i, normal_i} \odot \exp(-\lambda_g \nabla R_{low_i, normal_i})\|, \quad (2)$$

where ∇ denotes the gradient including ∇_x (horizontal) and ∇_y (vertical), λ_g represents structure perception coefficient. From Retinex theory, the illumination component becomes steep when it encounters the reflection component of the object, where

image structures locate and its illumination channel should be discontinuous. So, the weight $\exp(-\lambda_g \nabla R_{low_i, normal_i})$ denotes illumination component gradient becomes steep.

Considering the reflection component is determined by the nature of object, the reflection component should be consistent under any condition. So L_R can be defined as:

$$L_R = \sum_{i=1}^N \|R_{low_i} - R_{normal_i}\|. \quad (3)$$

Using reconstructed \tilde{S}_{low} and \tilde{I}_{low} , the reconstruction loss L_{Rec} can be formulated as:

$$L_{Rec} = \sum_{i=1}^N \|\tilde{R}_{low_i} * \tilde{I}_{low_i} - S_{normal_i}\|^2 \quad (4)$$

where $\tilde{R}_{low_i} * \tilde{I}_{low_i}$ denotes the enhanced image \tilde{S}_{low} . Here, with the pixel MSE loss function, the reconstructed \tilde{S}_{low} should be similar with its normal light version. But the loss function of the MSE optimization usually lacks the high-frequency content which results in perceptual unsatisfactory images with over-smooth texture. This is one reason why the reconstructed image is blurry.

To achieve more realistic reconstructed image, we introduce adversarial loss [18] to the objective loss, which is defined as follows:

$$L_{adv} = \frac{1}{N} \sum_{i=1}^N \log(1 - D_\theta(G_\omega(I_i^{LR}))), \quad (5)$$

where $D_\theta(G_\omega(I_i^{LR}))$ is probability that the reconstruction image $G_\omega(I_i^{LR})$ is a natural relighted image, and the function $G_\omega()$ means the output of the generator. The adversarial loss encourages the generator network to generate shaper high-frequency details. We combine the adversarial loss function with the pixel-wise MSE loss function for getting better natural reconstruction quality. So the generator network loss function L_G is defined as follows:

$$L_G = \frac{1}{N} \sum_{i=1}^N \log(1 - D_\theta(G_\omega(\tilde{R}_{low_i} * \tilde{I}_{low_i}))) + \left(\|G_\omega(\tilde{R}_{low_i} * \tilde{I}_{low_i}) - S_{normal_i}\|^2 \right) \quad (6)$$

where $D_\theta(G_\omega(\tilde{R}_{low_i} * \tilde{I}_{low_i}))$ denotes the probability that the reconstruction image $G_\omega(\tilde{R}_{low_i} * \tilde{I}_{low_i})$ is a natural normal light image S_{normal} . It enforces the reconstructed light images to be similar with the real ones on the pixel, high-frequency details and semantic level.

We also introduce classification loss to classify whether the reconstructed normal light images are fake or real. The loss function is defined as follows:

$$L_D = \frac{1}{N} \sum_{i=1}^N -(\log(1 - D_\theta(G_\omega(\tilde{R}_{low_i} * \tilde{I}_{low_i}))) + \log D_\theta(S_{normal_i})), \quad (7)$$

by introducing the classification loss, discriminator network makes images generated by generator more realistic. Therefore,

TABLE I
THE AVERAGE OF PSNR(DB), SSIM AND FSIM VALUES
WITH DIFFERENT LLIE METHODS

Methods	PSNR	SSIM	FSIM
PLE	13.99	0.7237	0.9123
SRIE	13.50	0.7148	0.9148
LIME	19.73	0.7598	0.8736
GladNet	18.26	0.7981	0.8596
RetinexNet	19.83	0.7031	0.8245
EnlightenGAN	20.06	0.8395	0.9214
Component-GAN	22.67	0.8570	0.9233

the GAN network can be trained by the objective function as:

$$L_{GAN} = \max_{\theta} \min_{\omega} \frac{1}{N} \sum_{i=1}^N \log(1 - D_\theta(G_\omega(\tilde{R}_{low_i} * \tilde{I}_{low_i}))) + \log D_\theta(S_{normal_i}) + \left(\|G_\omega(\tilde{R}_{low_i} * \tilde{I}_{low_i}) - S_{normal_i}\|^2 \right) \quad (8)$$

III. EXPERIMENTAL RESULTS

A. Datasets and Implementation Details

Component-GAN requires paired low/normal-light images for training. Considering it is difficult to collect paired low/normal-light images, we use the LOL dataset [12] and convert the normal-light image into a low-light version as training dataset. LOL is the first dataset containing low-light/normal image pairs from a real scene. It contains 500 pairs of low/normal-light images. We select 485 pairs as the training dataset, and the remaining 15 pairs as the testing dataset. To test the performance of Component-GAN, we also choose the standard images used in previous work as the testing datasets: (NPE [19], LIME [8], MEF [20], DICM [21], etc.). All samples are converted to PNG format and resized to 256×256 pixels. We use the random gradient descent method for back propagation and batch size is set to be 32. λ_I, λ_R and λ_g are set to be 0.1, 0.01 and 10 respectively.

B. Comparisons With Some State-of-The-Art Algorithms

In this section, we compare Component-GAN with current state-of-the-art methods including quantitative comparison, visual quality comparison, and no-referenced image quality assessment (IQA).

1) *Objective Evaluation:* In this subsection, we select three evaluation indexes: PSNR, SSIM, and FSIM to show the objective reconstruction quality. 15 pairs of low/normal-light images from LOL dataset are used as testing samples, in which low-light images are used as input and normal light images are used as groundtruth images. Six state-of-the-art LLIE methods: PLE [22], SRIE [7], LIME [8], GladNet [23], RetinexNet [12], EnlightenGAN [14] are selected as benchmarks for comparing. The average of indexes are listed in Table I.

2) *Visual Quality Comparison:* In order to show the visual results of Component-GAN, we list some reconstructed normal light images from the selected benchmarks. The results are demonstrated in Fig. 2 and Fig. 3, where the first column is the input low-light images, the second to seventh columns

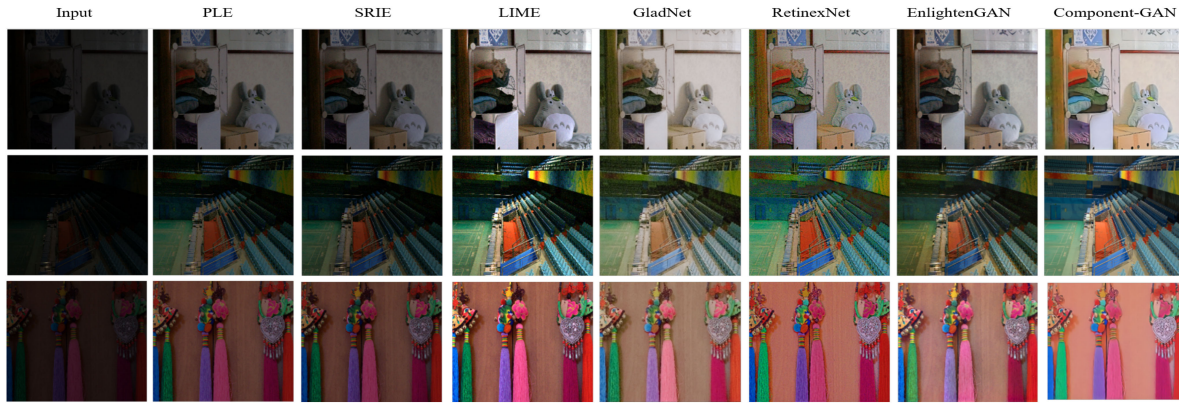


Fig. 2. Comparison with other state-of-the-art methods. We randomly selected three images from the LOL test dataset as input images, and reconstructed the results on the current excellent light enhancement algorithms.

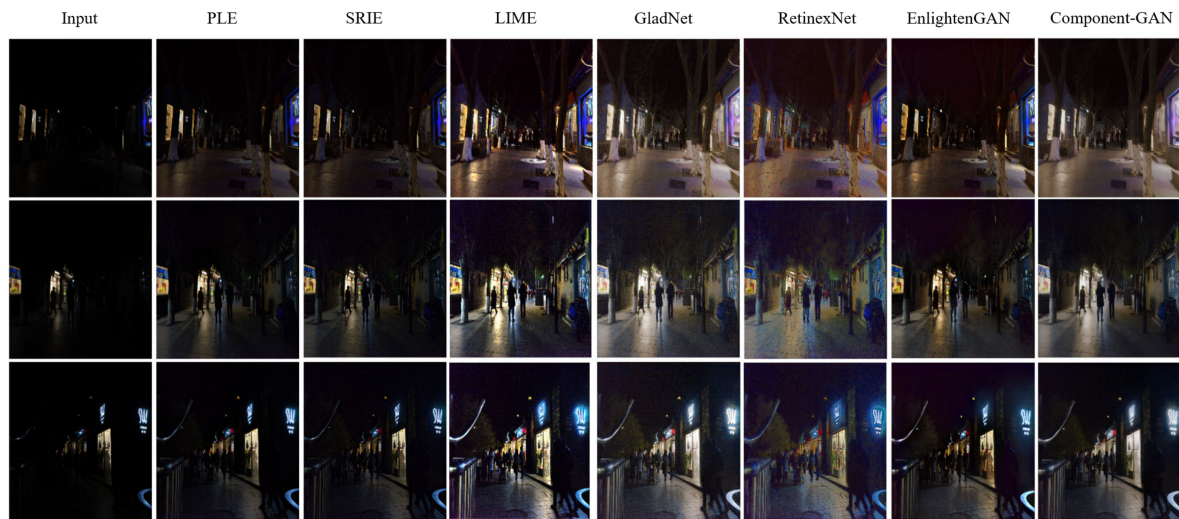


Fig. 3. Comparison with other state-of-the-art methods. We randomly selected three images as input images, and reconstructed the results on the current excellent light enhancement algorithms.

TABLE II
NIQE SCORES ON FIVE TESTING DATASETS (MEF, LIME, NPE, VV, DICM)

Image datasets	MEF	LIME	NPE	VV	DICM	All
Input testing	4.265	4.438	4.319	3.525	4.255	4.134
SRIE	3.475	3.788	3.986	2.850	3.899	3.650
LIME	3.720	4.155	4.268	2.489	3.846	3.692
RetinexNet	4.149	4.420	4.485	2.602	4.200	3.920
EnlightenGAN	3.232	3.719	4.113	2.581	3.570	3.385
Component-GAN	3.143	3.862	3.988	2.576	3.512	3.371

are the images enhanced by: PLE [22], SRIE [7], LIME [8], GladNet [23], RetinexNet [12], EnlightenGAN [14]. The last column shows the results produced by our Component-GAN.

In Fig. 2 and Fig. 3, the results from Component-GAN are brighter than EnlightenGAN. EnlightenGAN is an unsupervised learning method that does not require paired low/normal-light images training, thus its visual results are not as good as Component-GAN. The image reconstructed by Component-GAN is more real and clearer than the image reconstructed by GladNet and RetinexNet. The main reason is that GladNet and

RetinexNet lose high-frequency information during the process of enhancing light intensity, resulting in too smooth images.

3) *No-Referenced Image Quality Assessment*: Moreover, we adopt Natural Image Quality Evaluator (NIQE) [24], a well-known no-reference image quality assessment for evaluating real image restoration. The NIQE results on five publicly available image datasets used by previous works (MEF, NPE, LIME, VV, and DICM) are reported in Table II: a lower NIQE value indicates better visual quality. Specifically, MEF, NPE, LIME, VV and DICM contain 17,8,10,24 and 69 images respectively. Component-GAN wins two out of five and is the best in terms of overall averaged NIQE. This further testifies the superiority of Component-GAN in generating high-quality visual results.

IV. CONCLUSION

In this letter, we propose a novel low-light enhancement method named Component-GAN. By combining generative adversarial network with Retinex theory, the component enhancement networks do boost the LLIE performance. The convincing tests and satisfactory results have confirmed the effectiveness of Component-GAN. In the future, our work will focus on the super-resolution research of low-light images.

REFERENCES

- [1] F. Yu *et al.*, “Bdd100 k: A diverse driving video database with scalable annotation tooling,” 2018, *arXiv:1805.04687*.
- [2] E. H. Land, “The retinex theory of color vision,” *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, 1977.
- [3] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, “A multiscale retinex for bridging the gap between color images and the human observation of scenes,” *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.
- [4] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, “Efficient contrast enhancement using adaptive gamma correction with weighting distribution,” *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1032–1041, Mar. 2013.
- [5] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” in *Proc. Graph. Gems IV* 1994, pp. 474–485.
- [6] M. Kaur, J. Kaur, and J. Kaur, “Survey of contrast enhancement techniques based on histogram equalization,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 7, 2011.
- [7] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, “A weighted variational model for simultaneous reflectance and illumination estimation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2782–2790.
- [8] X. Guo, Y. Li, and H. Ling, “Lime: Low-light image enhancement via illumination map estimation,” *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.
- [9] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, “A fusion-based enhancing method for weakly illuminated images,” *Signal Process.*, vol. 129, pp. 82–96, 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0165168416300949>
- [10] J. S. Park and N. Cho, “Generation of high dynamic range illumination from a single image for the enhancement of undesirably illuminated images,” *Multimedia Tools Appl.*, vol. 78, no. 14, pp. 20263–20283, Jul. 2019.
- [11] M. Gharbi, J. Chen, J. Barron, S. Hasinoff, and F. Durand, “Deep bilateral learning for real-time image enhancement,” *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–12, Jul. 2017.
- [12] W. Y. J. L. Chen Wei, W. Wang, “Deep retinex decomposition for low-light enhancement,” in *Proc. Brit. Mach. Vis. Conf.*, 2018.
- [13] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, “Structure-revealing low-light image enhancement via robust retinex model,” *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.
- [14] Y. Jiang *et al.*, “Enlightengan: Deep light enhancement without paired supervision,” *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.
- [15] G. Chunle *et al.*, “Zero-reference deep curve estimation for low-light image enhancement,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1780–1789.
- [16] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. Med. Image Comput. Computer-assisted Interv.* 2015, pp. 234–241.
- [17] B. Goldluecke and D. Cremers, “An approach to vectorial total variation based on geometric measure theory,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 327–333.
- [18] C. Ledig *et al.*, “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proc. IEEE Conf. Computer Vis. Pattern Recognit.*, 2017, pp. 4681–4690.
- [19] S. Wang, J. Zheng, H.-M. Hu, and B. Li, “Naturalness preserved enhancement algorithm for non-uniform illumination images,” *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.
- [20] K. Ma, K. Zeng, and Z. Wang, “Perceptual quality assessment for multi-exposure image fusion,” *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
- [21] C. Lee, C. Lee, and C.-S. Kim, “Contrast enhancement based on layered difference representation,” in *Proc. 19th IEEE Int. Conf. Image Process.*, 2012, pp. 965–968.
- [22] X. Fu, Y. Liao, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, “A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation,” *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4965–4977, Dec. 2015.
- [23] W. Wang, C. Wei, W. Yang, and J. Liu, “Gladnet: Low-light enhancement network with global awareness,” in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2018, pp. 751–755.
- [24] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a “completely blind” image quality analyzer,” *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.