Bidirectional Convolutional-GRU Integrated GAN for Unpaired Low-Light Image Enhancement

Palak Deb Patra
Department of EE
NatWest Group
Bengaluru, Karnataka, India
palakdevpatra@gmail.com

Santosh Kumar Panda
School of Comp. Sci.
UPES Dehradun
Dehradun, Uttarakhand, India
sonupanda1995@gmail.com

Manoj Kumar Bishwal
Department of DASS
Siemens
Bengaluru, Karnataka, India
Manoj.bishwal@siemens.com

Pankaj Kumar Sa
Department of CSE
NIT Rourkela
Rourkela, Odisha, India
PankajKSa@nitrkl.ac.in

***Abstract*—Image enhancement is a widely researched area in the domain of computer vision, particularly image processing. Among the subdomains, low-light image enhancement (LLIE) receives considerable attention due to problems and challenges imposed by poor lighting conditions. As such, low-light images suffer from poor visibility, distorted colors, and loss of details, which limits their usability in many applications. The traditional methods have struggled to preserve such details and make the images susceptible to over-enhancement. Whereas, the learning-based techniques rely heavily on paired datasets for training. Therefore, we propose a Bidirectional Conv-GRU integrated GAN framework. Involving bidirectional Conv-GRU modules in our use-case enables the model to capture both local textures and long-range feature dependencies. Also, the use of unpaired datasets allows it to learn flexible and realistic mappings without the strict need for aligned image pairs. The results demonstrate the potency of our proposed work as compared to state-of-the-art methods.**

***Index Terms*– Low-light image enhancement, Generative Adversarial Networks (GAN), Bidirectional Conv-GRU, Unpaired dataset**

I. Introduction

LLIE is a crucial image preprocessing step in various applications, which include drone surveillance, night-time autonomous driving, and advanced medical imaging. Some traditional approaches, such as histogram equalization (HE) [1] and Retinex-based methods [2], have been thoroughly researched and implemented for this task. However, these techniques are quite stagnant in producing better results and may generate over-enhanced images. They primarily work by globally targetting all the pixels. Learning-based approach are also used for LLIE. They can produce very good enhancement results. Despite their given success, these methods are computationally expensive and rely on paired datasets for training. In practice, acquiring such paired low-light and well-lit bright images is quite hectic and non-ideal, leading researchers to generate their own synthetic paired datasets for training purposes.

GANs [3] emerged as a powerful framework in image generation and enhancement due to their ability to model complex data distributions. A GAN model consists of two network architectures: a generator (G), which generates fake images, and a discriminator (D), which distinguishes between real and generated fake samples. These networks are trained in an adversarial min–max game theory approach. According to the min-max game theory, the generator aims to produce outputs that are indistinguishable from real images. At the same time, the discriminator learns to classify and distinguish between real and fake images with utmost accuracy. Through this process, the generator improves its ability to generate realistic images. However, conventional GANs face challenges such as training instability, limited control over generated outputs, and issues like mode collapse, which have demanded the development of other improved variants of GAN [4], [5].

(a) Low-light (b) Groundtruth (c) Ours

Fig. 1: Example of an image enhanced using our propsoed work.

Traditional supervised GAN frameworks work on the concept of paired datasets. Such datasets are mostly difficult and impractical to obtain. Hence, we propose a model inspired by CycleGAN [6] that enables learning between domains without requiring paired data. This framework introduces a cycle consistency loss property, ensuring that an image is generated from one domain to the other and then reconstructed back, such that it's structure is consistent with the original. This property preserves the underlying content and structure while simultaneously adapting the image to the features of the target domain.

In our work, we have used Bidirectional Convolutional GRU (Conv-GRU) module inside the generator architecture of our model. Doing this provides an effective modelling of local spatial patterns as well as the long-range contextual dependencies. As seen in Fig. 1 our proposed work can enhance low-light image very efficiently.

The main objective of our work is to develop a new deep learning architecture for LLIE. The objectives are enumerated as follows:

1) A diverse unpaired dataset has been curated having low-light and bright-light images, which are independent of each other.
2) Develop a GAN-based framework that is highly capable of learning from unpaired datasets.
3) Integrate a Bidirectional Conv-GRU module inside the GAN generator.

The remaining parts of this paper are organized as follows. Section II gives an overview of related work in the field. Section III: Description of the proposed Bidirectional Conv-GRU integrated GAN framework. Section IV: Description of the experimental setup and discussions on the obtained results. Section V: Concluding the paper with key findings and some limitations.

II. Literature Review

Traditional methods such as HE [1] and Retinex-based methods [2] have laid the foundation for LLIE. Several variations of HE have been implemented by various researchers to improve quality of the image. Some other classical techniques include Fourier transformation-based methods, gamma correction, and logarithmic transformations. Researchers have also explored fusion-based strategies, such as multi-exposure [7] fusion, multi-sensor [8] image fusion, multi-scale [9], [10], [11] and ensemble-based enhancement methods, which combine outputs from multiple models to improve quality.

But one major change started after the growth of deep learning, which significantly shifted LLIE research. One of the early works in this direction was LLNet, proposed by Lore *et al.* [12] in 2017, which proposed a stacked sparse autoencoder to jointly enhance contrast and suppress noise. Building on this line of work, Deep Retinex Decomposition [13] proposed a deep learning framework by separating an image into illumination and reflectance components. With the due course of time, a wide range of approaches based on graph convolution, zero-shot learning strategies [14], and more recent developments, such as Mamba-based architectures [15] aimed at efficient modelling for LLIE tasks.

But it was the CycleGAN that made it possible to translate images from one to another without pairing them. DualGAN [16], which used dual learning with paired generators and discriminators, was effective in image translation without the need for paired datasets.

Some more recent works have explored the integration of recurrent units, particularly gated recurrent units (GRUs) [17], to enhance contextual modelling in LLIE. Dou *et al.* [18] proposed a multi branch iterative network based on GRU for LLIE.

III. Methodology

A. Dataset Collection

Our study uses a diverse dataset compiled from multiple publicly available sources like LOL [19] and SICE [20]. There are two independent subsets, referred to as Dataset X and Dataset Y. Dataset X consists of low-light images, whereas

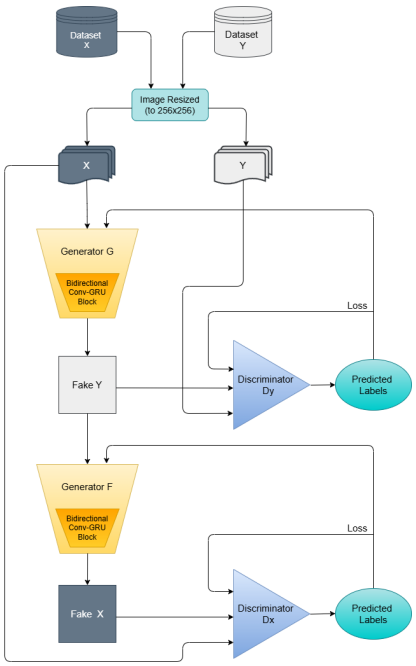

Fig. 2: Overall architecture of our model

Dataset Y contains normal-light (or well-lit) images. It may be noted that these two dataset pairs are completely independent of each other, making them a suitable candidate for unpaired training. These subsets are treated as unpaired data, allowing the model to learn mappings between both domains. All images are resized to 256×256 pixels. Resizing the images to a lower resolution, helped us in easier computations and also introduced uniformity to the input images that had variable resolutions.

B. Generators

Our work comprises two generators as shown in the overall architecture in Fig. 2, to enable bidirectional translation between domains:

- **Generator G**: Translates low-light images from domain X to bright-light images in domain Y.
- **Generator F**: Maps well-lit images from domain Y back to low-light images in domain X.

Each generator as shown in Fig. 3, follows the principle of an encoder–decoder architecture. The main component or the backbone of the architecture is the BiConv-GRU, which integrates recurrent memory with convolutional operations. Through these operations it capture spatial details and long-range contextual dependencies very efficiently. As we can see, the overall architecture is divided into three main stages:

1) Encoding (Downsampling): The input image of size 256×256 is processed through a series of convolutional layers with increasing channel depth and downsampling operations. This stage extracts low-level textures and high-level semantic features.

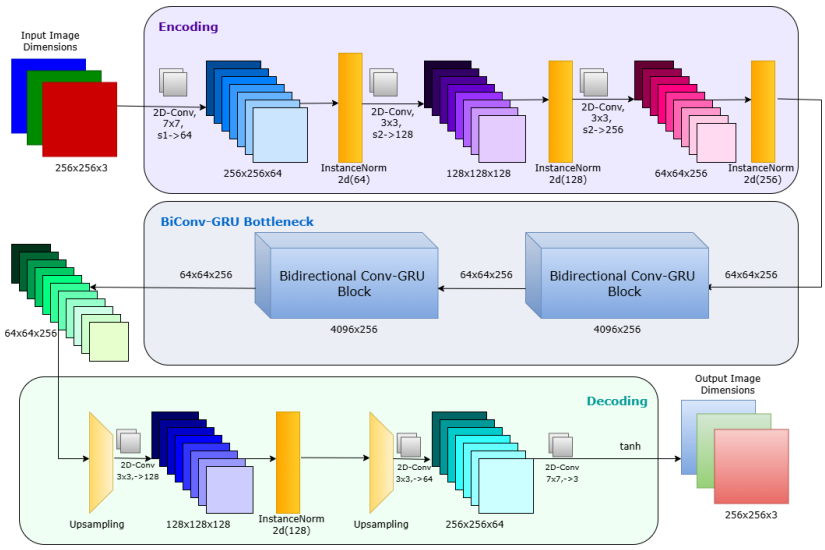

Fig. 3: Architecture of Generator

2) Bottleneck (BiConv-GRU Blocks): At the bottleneck, two BiConv-GRU blocks are utilized to refine the encoded feature representations. These modules operate bidirectionally, which means the model captures contextual dependencies in both directions. This enhances the model's ability to restore the regions which are underexposed while also maintaining structural consistency.

3) Decoding (Upsampling): Finally, in this stage, the refined feature maps are passed through convolutional layers to restore the original resolution. This stage basically reconstructs the enhanced image.

Overall, this inclusion of BiConv-GRU modules ensures robust contextual modelling, while the symmetric encoder–decoder structure contributes better to training and reconstruction.

C. Bidirectional Conv-GRU

BiConv-GRU blocks constitute the core part of the generator, combining convolutional operations with bidirectional recurrent processing to effectively capture both fine-grained spatial details and long-range contextual dependencies. Each block is composed of the following components:

1) Depthwise 1-D Convolution: Extracts localized spatial features by capturing interactions among neighboring elements within the feature map while maintaining computational efficiency.

2) Gating Mechanism: Introduces non-linearity and adaptive feature modulation through element-wise activation, defined as $x = x \cdot \sigma(x)$, enabling selective feature propagation.

3) Forward GRU: Processes features sequentially in the forward direction (e.g., left to right), capturing contextual dependencies from preceding spatial locations.

4) Backward GRU: Processes features in the reverse direction (e.g., right to left), capturing complementary contextual information from subsequent spatial locations.

5) Feature Fusion with Residual Connection: Integrates forward and backward representations along with the original input features, ensuring information preservation and stable gradient flow.

6) Layer Normalization: Normalizes intermediate activations to improve training stability and convergence.

7) Linear Projection: Adjusts feature dimensionality to facilitate seamless integration with subsequent network layers.

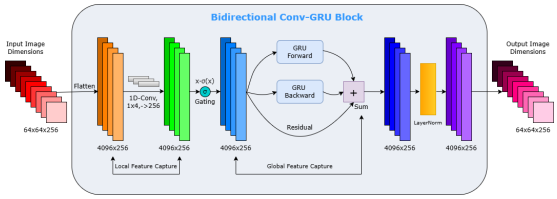

Fig. 4: Architecture of BiConv-GRU Block

D. Discriminators

The proposed GAN framework incorporates two discriminators as shown in Fig. 5, D_X and D_Y, each designed to evaluate image realism at the patch level rather than over the entire image. This patch-based discrimination enables the model to focus on high-frequency details and local texture consistency.

- **Discriminator** D_X assesses the authenticity of low-light images by distinguishing real samples from domain X and generated samples obtained through the mapping $F : Y \rightarrow X$.
- **Discriminator** D_Y performs a similar function for well-lit images, differentiating real samples from domain Y and synthesized outputs generated via $G : X \rightarrow Y$.

Both discriminators are implemented using a PatchGAN architecture, as illustrated in Fig. 4. Instead of producing a single realism score for the entire image, PatchGAN evaluates

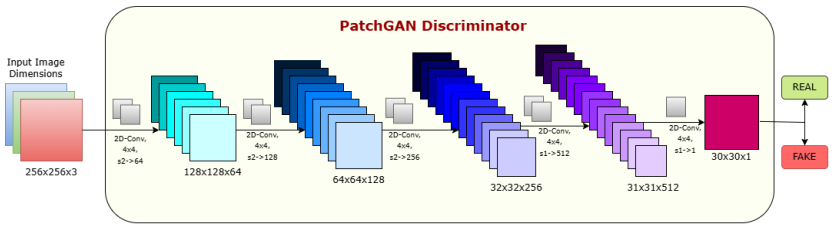

Fig. 5: Architecture of Discriminator

overlapping local patches and classifies each patch as real or fake, resulting in a matrix of output responses. Each element in this output map corresponds to a receptive field of size 70×70 in the input image.

IV. Experiment and Results

A. Training Parameters

The proposed CycleGAN was trained for 150 epochs with model checkpoints saved every 5 epochs. Training used unpaired samples from domains X (low-light) and Y (bright-light). Models were trained on mini-batches with standard adversarial updates for generators and discriminators. The experiment was conducted on a NVIDIA RTX 5070 ti OC GPU system with 16GB VRAM, and the model was implemented using Pytorch framework.

B. Optimizers and Schedulers

Adam optimizer was used for both generators and discriminators with the following hyperparameters [6]:

- Generator optimizer: Adam($\theta_{G,F}$, lr $= 2 \times 10^{-4}$, $\beta = (0.5, 0.999)$).
- Discriminator optimizers: Adam(θ_{D_X,D_Y}, lr $= 2 \times 10^{-4}$, $\beta = (0.5, 0.999)$).

Learning rates were controlled by CosineAnnealingLR schedulers (one per optimizer) with $T_{\max} = 150$. Learning rates were updated using an epoch-wise scheduling strategy

C. Loss Functions

Training minimized a composite objective containing adversarial, cycle-consistency, identity and perceptual terms. Weights used in training were chosen empirically and held fixed:

$$\mathrm{W}_{\text{adv}} = 1.0, \quad \mathrm{W}_{\text{id}} = 5.0, \quad \mathrm{W}_{\text{cyc}} = 10.0, \quad \mathrm{W}_{\text{perc}} = 1.0.$$

1) Adversarial Loss: Adversarial loss trains the generators and discriminators using a least-squares approach, implemented with an MSE loss, and is responsible for making the generated images appear realistic to the discriminators.

$$\mathcal{L}_{\text{adv}}(G, D_Y) = \mathbb{E}_y \left[(D_Y(y) - 1)^2\right] + \mathbb{E}_x \left[D_Y(G(x))^2\right] \quad (1)$$

$$\mathcal{L}_{\text{adv}}(F, D_X) = \mathbb{E}_x \left[(D_x(x) - 1)^2\right] + \mathbb{E}_y \left[D_x(F(y))^2\right] \quad (2)$$

2) Cycle-Consistency Loss: Cycle-consistency loss ensures that an image mapped to the other domain and then returns back to its original form. It is implemented using an L1 loss:

$$\mathcal{L}_{\text{cyc}} = \mathbb{E}_x \|F(G(x)) - x\|_1 + \mathbb{E}_y \|G(F(y)) - y\|_1 \quad (3)$$

3) Identity Loss: Identity loss discourages the generator from altering images that already belong to the target domain and is implemented using an L1 loss.

$$\mathcal{L}_{\text{id}} = \mathbb{E}_x \|F(x) - x\|_1 + \mathbb{E}_y \|G(y) - y\|_1, \quad (4)$$

4) Perceptual loss: A VGG-based perceptual term improves high-level feature consistency. VGG16 features up to `conv3_3` (layer 16) are extracted from a frozen VGG16. Perceptual loss is the ℓ_1 distance between VGG feature maps of the generated and ground-truth images:

$$\mathcal{L}_{\text{perc}} = \mathbb{E} \|\phi(\tilde{x}) - \phi(x)\|_1 + \mathbb{E} \|\phi(\tilde{y}) - \phi(y)\|_1, \quad (5)$$

where $\phi(\cdot)$ is the truncated VGG16 feature extractor.
The total generator loss minimized in each update step is:

$$\begin{aligned} \mathcal{L}_G^{\text{total}} = \; & \mathrm{W}_{\text{adv}} \cdot \mathcal{L}_{\text{adv}}(G, D_Y) \\ & + \mathrm{W}_{\text{adv}} \cdot \mathcal{L}_{\text{adv}}(F, D_X) \\ & + \mathrm{W}_{\text{cyc}} \cdot \mathcal{L}_{\text{cyc}} \\ & + \mathrm{W}_{\text{id}} \cdot \mathcal{L}_{\text{id}} \\ & + \mathrm{W}_{\text{perc}} \cdot \mathcal{L}_{\text{perc}} \end{aligned} \quad (6)$$

D. Results

The proposed model was evaluated using many objective evaluation metrics like PSNR, SSIM, and MSE. A subjective evaluation is also done where the proposed approach is compared with SOTA methods like RetinexNet [19], SGM [21], DRBN [22], DSLR [23], EnlightenGAN [24], UretinexNet [25], and ZeroDCE [14], and the results are shown in Fig. 6. From the figure it can be seen that many of the SOTA methods performed poorly. Some of the images are washed out whereas some images have lost important features of the original groundtruth image. An objective evaluation has also been performed and shown in Table. I where the models are compared based on their PSNR, MSE and SSIM values. The arrow keys represent the significance of their values. $\uparrow$ means higher the value the better and $\downarrow$ means lower the value the better. It can be seen clearly from the table that our proposed approach has better results than other benchmark methods. This performance improvement between our method and other methods can be attributed to the use of

TABLE I: Objective evaluation of our proposed model as compared with SOTA methods.

	RetinexNet	SGM	DRBN	DSLR	EnlightenGAN	UretinexNet	ZeroDCE	Proposed
PSNR ↑	14.19	14.82	13.86	12.54	14.95	17.88	13.51	**21.01**
MSE ↓	2018.63	2064.56	3256.23	3577.85	2392.84	1169.24	2834.78	**554.25**
SSIM ↑	0.64	0.63	0.68	0.71	0.71	**0.74**	0.64	0.72

Bidirectional Conv-GRU module used in the generator that effectively captures both spatial and long range contextual dependencies.

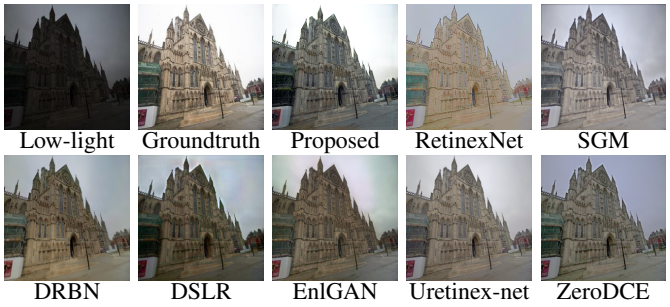

Fig. 6: Subjective evaluation of proposed method on an image from SICE dataset.

V. Conclusion

In this study, we propose a novel integration of Bidirectional Convolutional Gated Recurrent Units (BiConv-GRU) within a GAN framework to address the problem of LLIE. Incorporating recurrent spatial modelling enabled the model to capture both local and global dependencies, resulting in improved enhancement. Experimental results demonstrates consistent improvements in illumination correction, structural fidelity, and perceptual quality, underscoring the effectiveness of the proposed architecture for real-world applications. However, certain limitations still exist. Since the dataset used for training was unpaired, pixel-level metrics did not reach the ideal values typically observed in paired supervision scenarios. While the model achieved good performance, some minor residual issues such as color shifts and incomplete fine-detail recovery remain evident in few cases. In future these limitations can be solved by a more robust training approach and hyperparameter fine tuning.

References

[1] R. C. Gonzalez, *Digital Image Processing*. Prentice Hall, 1992.

[2] E. H. Land, "The retinex theory of color vision," *Scientific American*, vol. 237, no. 6, pp. 108–129, 1977.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[4] A. Patil *et al.*, "Dcgan: Deep convolutional gan with attention module for remote view classification," in *2021 International Conference on Forensics, Analytics, Big Data, Security (FABS)*, vol. 1. IEEE, 2021, pp. 1–10.

[5] J. Adler and S. Lunz, "Banach wasserstein gan," *Advances in neural information processing systems*, vol. 31, 2018.

[6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[7] S. K. Panda and P. K. Sa, "Exposure calibration network with graph convolution for low-light image enhancement," *IEEE Sensors Journal*, vol. 24, no. 24, pp. 41 971–41 980, 2024.

[8] Panda, Santosh Kumar and Sa, Pankaj Kumar, "A thermal-aware multi-sensor framework for low-light image enhancement," *IEEE Sensors Journal*, 2025.

[9] S. Das, M. Roy, and S. Mukhopadhyay, "Correcting low illumination images using pso-based gamma correction and image classifying method," in *International Conference on Computer Vision and Image Processing*. Springer, 2020, pp. 433–444.

[10] Das, Swadhin and Roy, Manali and Mukhopadhyay, Susanta, "Correcting low-illumination images using multi-scale fusion in a pyramidal framework," in *2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*. IEEE, 2020, pp. 126–129.

[11] S. K. Panda, D. Nayak, and P. K. Sa, "Mccnet: A multi-scale cross connection network for low-light image enhancement," in *International Conference on Computer Vision and Image Processing*. Springer, 2023, pp. 488–499.

[12] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.

[13] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.

[14] 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 *Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 1780–1789.

[15] R. Deng, A. Jiang, L. Peng, and Q. Yan, "Codebook knowledge with mamba-transformer for low-light image enhancement," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2026, pp. 3720–3729.

[16] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2849–2857.

[17] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (gru) neural networks," in *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*. IEEE, 2017, pp. 1597–1600.

[18] Y. Dou, Y. Gao, M. Gao, S. Zhao, and C. Zeng, "Multi-branch low-light image iterative enhancement network," *Scientific Reports*, vol. 15, no. 1, p. 43078, 2025.

[19] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *British Machine Vision Conference*, 2018.

[20] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2049–2062, 2018.

[21] W. Yang, W. Wang, H. Huang, S. Wang, and J. Liu, "Sparse gradient regularized deep retinex network for robust low-light image enhancement," *IEEE Transactions on Image Processing*, vol. 30, pp. 2072–2086, 2021.

[22] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, "Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality," *IEEE Transactions on Image Processing*, vol. 30, pp. 3461–3473, 2021.

[23] S. Lim and W. Kim, "DSLR: Deep stacked Laplacian restorer for low-light image enhancement," *IEEE Transactions on Multimedia*, vol. 23, pp. 4272–4284, 2020.

[24] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.

[25] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement," in *Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5901–5910.