World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:18, No:11, 2024

Low Light Image Enhancement with Multi-Stage Interconnected Autoencoders Integration in Pix-to-Pix GAN

Muhammad Atif, Cang Yan

***Abstract*—**The enhancement of low-light images is a significant area of study aimed at enhancing the quality of captured images in challenging lighting environments. Recently, methods based on Convolutional Neural Networks (CNN) have gained prominence as they offer state-of-the-art performance. However, many approaches based on CNN rely on increasing the size and complexity of the neural network. In this study, we propose an alternative method for improving low-light images using an Autoencoders-based multiscale knowledge transfer model. Our method leverages the power of three autoencoders, where the encoders of the first two autoencoders are directly connected to the decoder of the third autoencoder. Additionally, the decoder of the first two autoencoders is connected to the encoder of the third autoencoder. This architecture enables effective knowledge transfer, allowing the third autoencoder to learn and benefit from the enhanced knowledge extracted by the first two autoencoders. We further integrate the proposed model into the Pix-to-Pix GAN framework. By integrating our proposed model as the generator in the GAN framework, we aim to produce enhanced images that not only exhibit improved visual quality but also possess a more authentic and realistic appearance. These experimental results, both qualitative and quantitative, show that our method is better than the state-of-the-art methodologies.

***Keywords*—**Low light image enhancement, deep learning, convolutional neural network, image processing.

I. Introduction

THE optimization of imagery obtained under low-illumination conditions has long been a focus of ongoing research. Such efforts hold valuable applications in nighttime monitoring systems, autonomous platforms, and digital image editing tools. When light levels are inadequate, captured scenes may lack visual acuity and fine element detail. This insufficiency challenges the interpretability of resulting images and influences downstream evaluations of visual information. Concurrently, poor illumination can impact assessments of higher-order scene understanding, like segmentation of scene [1], [2] and face-recognition [3], [4].

Two primary conventional strategies for improving images under low light condition are techniques based on histogram equalization and Retinex [5], [6]. The Histogram equalization is a technique that improves the overall levels of lighting in an image by shifting how frequently pixels with varying brightness values occur in the frequency distribution. Retinex-based techniques [7], [8] split the input picture in two components one is reflection component and illumination factor and then modify the illumination factor or utilize reflectance straight to get an improved output. Conventional approaches leveraging prior information have demonstrated capable performance. Nevertheless, traditionally prior-driven approaches are unable to take into consideration a wide range of lighting conditions and the outcomes that are produced frequently conceal specifics and maintain uneven illumination. Data driven algorithms are now widely used in computer vision because of deep learning's potent feature representation capabilities. The majority of low-light image denoising models are based on CNN, to match the mapping association from a large amount of data between low light and normal light images. Deep learning method to improve images are LLNet [9], the first CNN-based method that simultaneously learns low-light picture enhancement and denoising using a sparse denoising autoencoder. Many deep learning-based techniques have since been presented, including RetinexNet [10], MBLLEN [11], Knowledge distillation [12], LACN [13], DTSD [14], EnlightenGAN [15], all of which have significant advancement and produced striking visual outcomes. Unfortunately, most algorithms require more memory and computational power to run well, making them unsuitable for real world applications like edge computing and mobile platforms. This is because most algorithms increase computation and parameter counts while optimizing algorithm performance. The autoencoder [16] is used to learn effective data encodings in unsupervised manner. Autoencoders use an encoder-decoder architecture to learn distributed, compressed representations of data. The decoder reconstructs the input from the lower-dimensional latent space representation that the encoder has learned. A dual autoencoder network [17] has purpose to improving low light images. In the past, Generative Adversarial Networks (GANs) [18] have shown encouraging performance in computer vision. A unique unsupervised method for improving low-light photos is suggested by the LE-GAN network [19], which addresses difficulties with noise, color bias, and over-exposure. It enhances visual quality without requiring paired training data by augmenting feature extraction with an illumination-aware attention module.

For low-light image enhancement, we provide a multistage deep learning architecture. The first two autoencoders function to provide an initial enhancement, with the encoder of first two autoencoder transferring its learned feature representation to the

Muhammad Atif is pursuing master degree Information and Communication Engineering, Harbin Engineering University, Harbin, China (phone: 86-18846107185; e-mail: Muhammadatif@hrbeu.edu.cn).

Cang Yan is the Professor in Department of Information and Communication Engineering, Harbin Engineering University, Harbin, China (phone: 86-1376682806; e-mail: cangyan@hrbeu.edu.cn).

decoder of the third autoencoder.

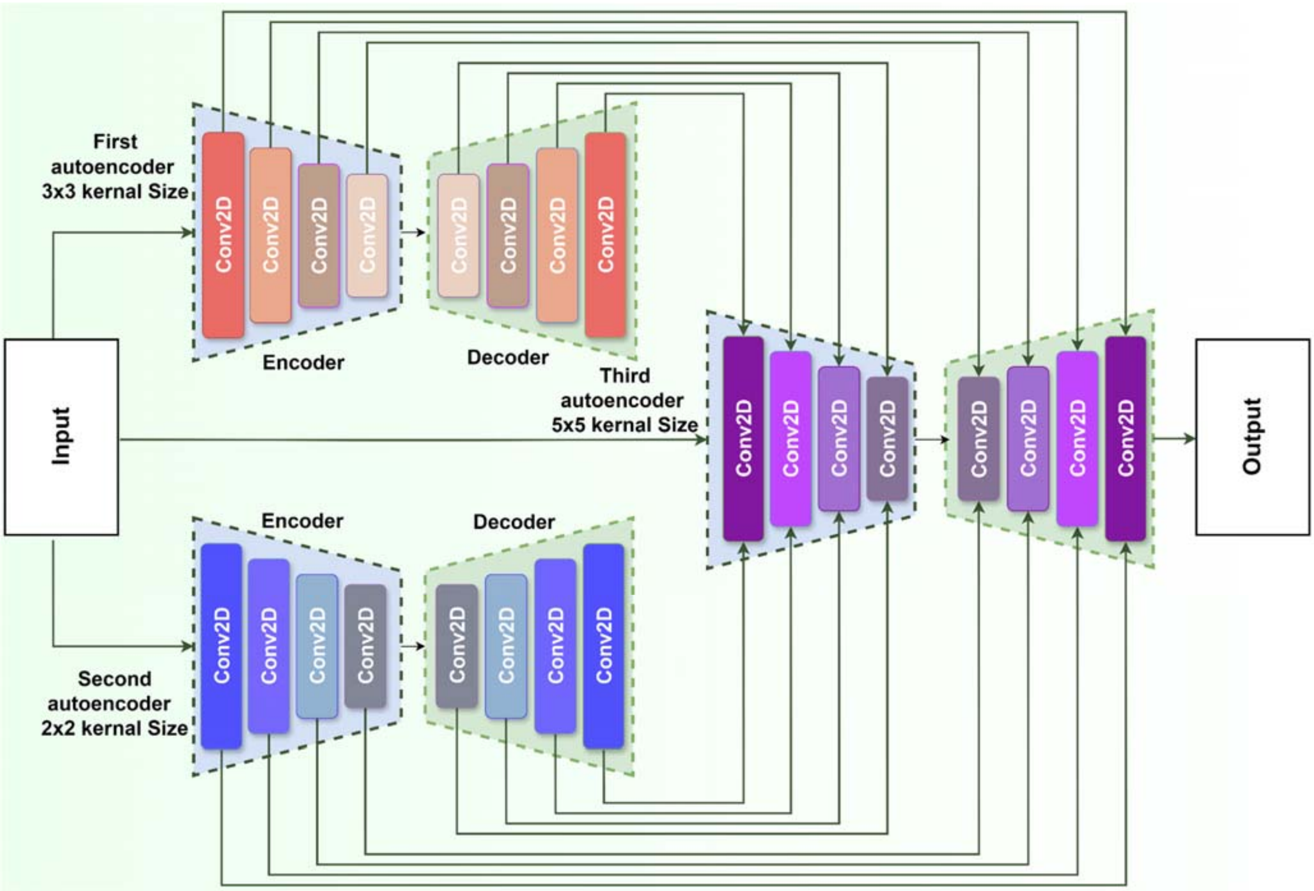

Fig. 1 The architecture of Generator Network for Low-Light Image Enhancement

As shown in Fig. 1, the decoder part of the first autoencoder and second autoencoder reconstructed output is then provided to a third autoencoder, designed to furnish a higher-fidelity enhancement. In this final autoencoder, knowledge is shared bi-directionally between its encoders and decoder segments. Specifically, the decoders inform the encoders as to what patterns and textures were effectively captured from the prior autoencoder's output, while the encoders concurrently guide the decoders on reconstructing finer details. This cooperative training scheme enables more sophisticated reconstruction capable of resolving subtle color and lighting variations. The proposed model is integrated into the Pix-to-Pix GAN framework. Given Pix-to-Pix employs conditional adversarial learning to complete the enhancement pipeline, generating results with high perceptual quality and naturalness. Through this multi-pronged approach fusing the self-supervised feature disentangling of autoencoders with Pix-to-Pix is supervised feature generation, our goal is to maximize both subjective and objective quality measures for photos taken in low-light environment.

This is a summary of the main contributions made by this work:

- We present a multi-stage deep learning architecture that combines multiple autoencoders with Pix-to-Pix GAN for the purpose of low light image enhancement. This phased strategy makes use of each method's supervised and unsupervised modeling capabilities.
- Our model utilizes a cascaded autoencoder structure where encoders and decoders share knowledge bidirectional to iteratively improve the enhanced output. This cooperative training scheme captures fine-grained image pattern.
- The autoencoder reconstructions are fed into Pix-to-Pix GAN for photo-realistic image translation to the high-light domain. This completes the enhancement with state-of-the-art perceptual quality.
- We designed a generator loss by combining the MAE loss, SSIM loss and Carbonnier loss.
- Extensive experiment on benchmark datasets demonstrates that, in terms of standard image quality, the suggested method outperforms alternative strategies in terms of both qualitative and quantitative enhancement and generation performance.

II. Related Work

A. Traditional Enhancement Method

For a long time, researchers have been actively investigating methods for improving low-light images, including Retinex modeling and histogram equalization. In order to address the issue of noise amplification, [5] suggested limiting the maximum amount of contrast amplification and performing histogram equalization on certain areas rather than the entire image. A formulation of the cumulative function was proposed by [20] to generate a mapping from a local histogram of the grayscale levels of pixels. The study of [21] demonstrated grayscale differences using a two-dimensional histogram organized in a hierarchical tree-like manner, grayscale differences should be more apparent in high-frequency zones. The problem is converted into K-dimensional space to create a reversible cumulative probability distribution function and calculate a strict pixel ordering for the image [22].

Drawing from the concepts of color constancy, the Retinex theory [7] enables decomposing the image into a reflection factor and an illumination component, where the illumination factor depicts the light incident on the item, and the reflection factor characterizes the object's intrinsic qualities. In the algorithmic domain, SSR [23] suggested Gaussian surround functions to solve for the illumination component and reflectance component, with the reflection factor being used as the improved result. In order to process the original image, a multiscale Retinex model using different sized Gaussian filters is presented by MSR [24]. Guo et al. [25], [26], using the enhanced Lagrange multiplier method, optimize the illumination factor by adding structural smoothing to the prior data and starting with the maximum pixel value. In order to perform Retinex decomposition [27], a bright-pass filter specifically suited for non-uniform pictures, using a double logarithmic transformation to maintain naturalness and details. To estimate the noise map [28], a noise component is added to the original Retinex model, putting restrictions on the segmental smoothness of the illumination map and the gradient consistency of the reflectance map. The reflectance and illumination components are thoroughly decomposed using a morphological closed-loop operation [29]. The two illumination maps are then combined using a weighted fusion approach that was produced through curve modification and histogram equalization. A weighted variational model was proposed as a means of improving the estimation of illumination factor and reflectance factor [8]. However, the capacity to recover details is restricted and approaches based on the Retinex model and histogram equalization frequently lose details in image and blurry edge features.

B. Deep Learning Based Methods

The approaches of deep learning have been effectively used in the area of low-light image enhancement and have gained prominence within related study due to the explosive progress of deep learning technology in the area of computer vision. The use of deep learning technology for low-light image enhancement, [9] suggested using an autoencoder structure for both denoising and low-light image enhancement simultaneously. Retinex theory was the foundation [10] which broke down the light and reflectance maps. In order to produce the output image; [11] suggested using a multi-branches enhancing fusion. In order to enhance images, [30] suggested a network for global light perception and details retention. A full-resolution illumination map is obtained [31] by first extracting global and local information at poor quality of resolution, followed by bilateral grid-based Upsampling. The study of [32] first utilizes a decomposition network to split the original image into illumination factor and reflectance factor. After that, they employ a reflectance map recovery network and an adjustment network, respectively, to alter the illumination factor and reflectance factor. In order to improve the image captured under low-light condition, the light back-projection was proposed [33] which also constructed the deep lightening network in a cascading manner. They then added an adjustable light control factor parameter. In order to guarantee temporal consistency in low-light video enhancement, [34] suggested simulating video from a single using an optical flow estimating method. Based on prior understanding of Retinex theory, [35] suggested a neural network architectural search method to automatically find effective network architectures from a predetermined search area. In order to prevent color bias, [36] suggested the Retinex-based self-regularized approach and carried out recovery in HSV color space. The study of [37] involves training the network through a series of unsupervised losses and relies on high-order curve adjustment in the absence of ground truth images. A semi-supervised method of picture enhancement [38] which uses unsupervised learning to raise the perceived images quality and picture fidelity restoration by supervised learning.

C. Generative Adversarial Network Methods

The application of GANs to deep learning approaches for low-light picture enhancement has demonstrated potential in terms of learning the mapping between low-light and normal-light image domain. A GAN-based domain adaptation strategy for learning from paired and unpaired data [39]. The study [15] suggested GANs for unpaired learning with global and local discriminators to improve low-light photos. The perceptual-details GAN [40] employs fractional differential gradient masks in the discriminator to suppress noise and enhance details while recovering illumination from extremely noisy low-light images. It does this by combining a Zero-DCE model and residual dense-block autoencoder. A progressive GAN-based transfer network that makes use of Retinex theory [41] is an efficient way to improve low-light photos. With the capacity to learn both the mapping and the loss function, the study of [42] presented a potent solution for image-to-image translation tasks. This method has been effectively used for a number of tasks, including object recognition and photo synthesis.

III. Methodology

This section first provides details of the proposed models including Generator and Discriminator model and an overview of the PIX2PIX GAN framework and their training. Afterwards, we go over the loss functions used in optimization, including SSIM, MAE, Generator, and Discriminator losses.

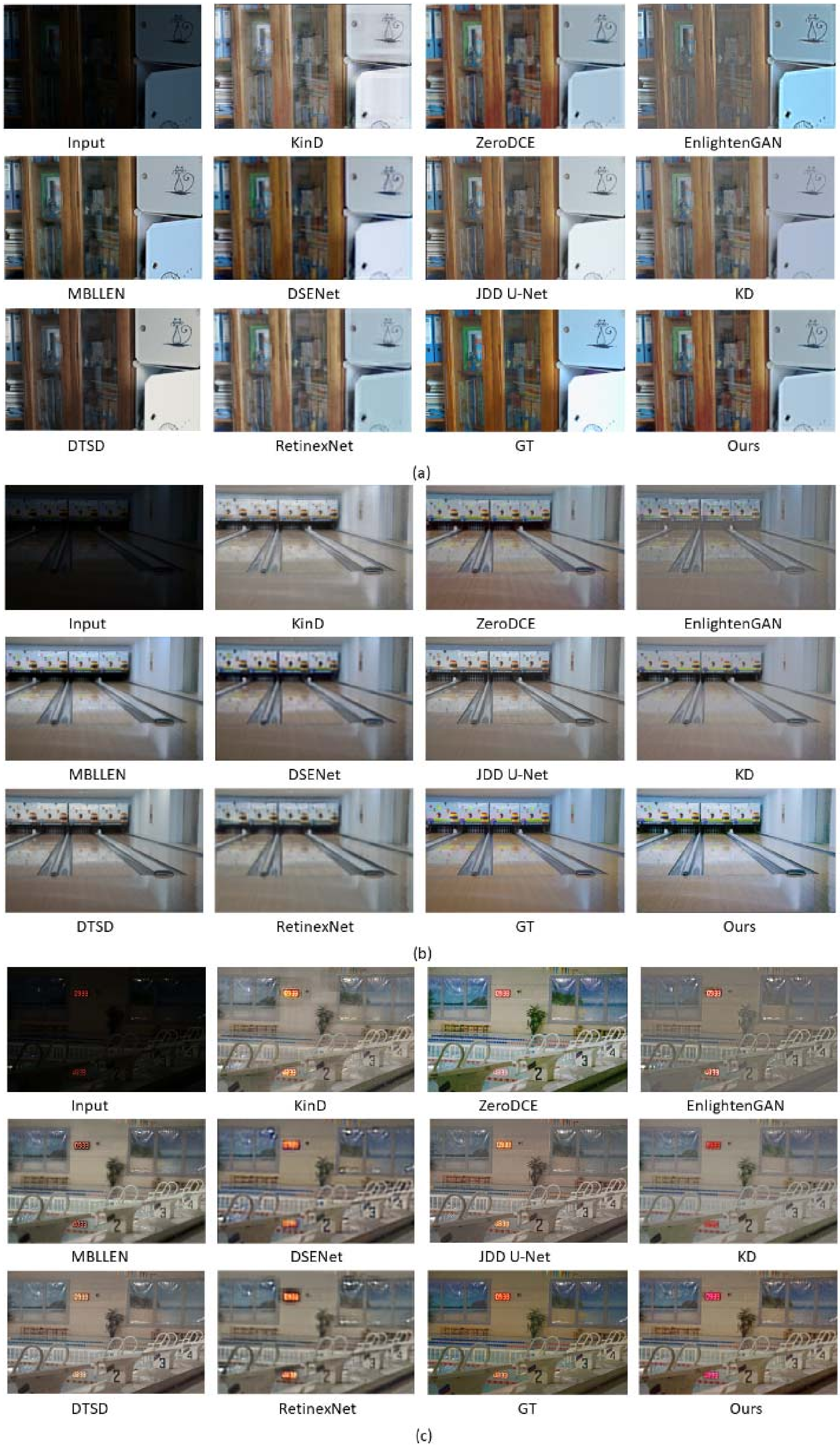

Fig. 2 Qualitative visual comparison outcomes of the various techniques using the LOL dataset

A. Network Architecture

1) Generator

We present a network architecture, as shown in Fig. 1 that leverages the power of autoencoders to enhance low-light image. Our network comprises three interconnected autoencoders: A, B and C. Autoencoder A and B are specifically designed to process low light images as input, the encoder sections of both autoencoder A and B progressively reduce the dimensions of the input image through multiple layers. Notably, each layer in encoder A and encoder B is connected to a corresponding layer with the same dimensions in the decoder of autoencoder C. This interconnected design enables the transfer of knowledge from autoencoders A and B to autoencoder C at every scale of the decoder. By establishing these connections, our network aims to capture and integrate more comprehensive information from the input image. In doing so, it facilitates the generation of high-quality outputs, effectively enhancing the visual appeal of photos taken in poor light.

Furthermore, we extend the knowledge transfer by connecting the decoders of autoencoders A and B to the corresponding layers in the encoder of autoencoder C. This connection allows the improved quality of the low light image obtained from autoencoder A and B to be transmitted to the encoder of autoencoder C. This transfer of knowledge helps enhance the encoder of autoencoder C further, leveraging the improved features extracted from the input images.

By exploiting the hierarchical representation learned by cross interconnected autoencoders, the network effectively combines the benefits of multi-scale information extraction and reconstruction. By using this method, the network is able to enhance the overall visual quality of the low-light photos by utilizing the rich contextual information that the encoders have acquired.

2) Discriminator

The discriminator architecture is based on a convolutional neural network. Regarding the challenges involving image-to-image translation, the discriminator acts as a binary classifier that is crucial for the GAN framework to discern between real and generated ones. Specifically, the discriminator contains convolutional block sequence that gradually downsample the input to extract multi-scale hierarchical features. As shown in Fig. 3, each convolutional block comprises of convolution-BatchNorm-LeakyReLU layers to apply linear transformations, normalize activations and introduce non-linearities. Global average pooling is used to combine feature maps into a single vector that represents the image after down-sampling. This is allowed by additional convolutional layers to synthesize hierarchical representations into a final binary classification. Through this architecture, the discriminator is able to capture both local textures and global structures to make a holistic judgment on image realism. The extracted hierarchical features are then used to calculated adversarial losses to provide training signals for the generator to match the distribution of data for actual images.

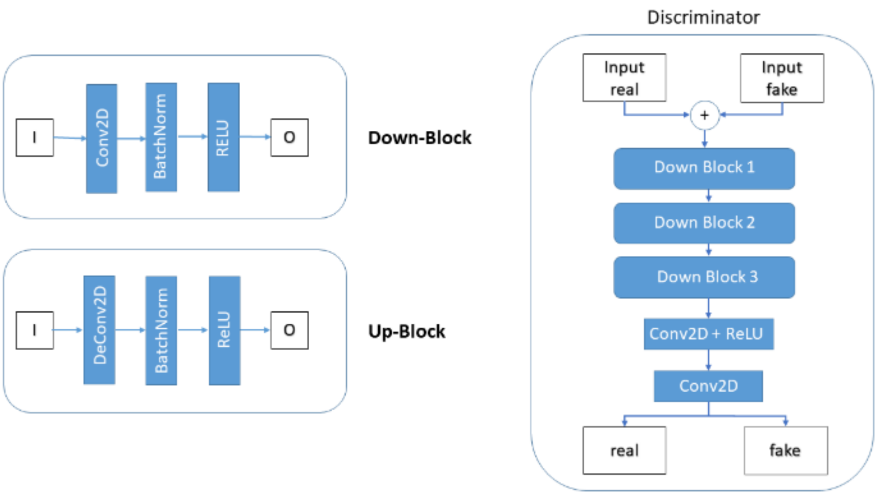

Fig. 3 The architecture of discriminator network

B. Generative Adversarial Networks

GANs discover how to translate a random noise vector z to an output image y, denoted as G:z→y. The Generator takes the noise vector z as input and produces an image y as the output without any conditioning information. Conditional GANs, on other hand, learn a mapping G:{x,z}→y, between an observed picture x and a random noise vector z to an output image y. The generator in a cGAN creates an output picture y that is conditioned on both the observed image x and the noise vector z as input. Discriminator Network D is trained to discriminate between actual photos and generate images as "fake", while Generator Network G is trained to produce outputs that are indistinguishable from real images, looking realistic and capable of "fooling" discriminator. The conditional GAN's objective can be described as:

$$L_{cGAN}(G, D) = E_{x,y}[\log D\,(x, y)] + E_{x,z}[\log(1 - D(x, G\,(x, y))] \tag{1}$$

where D tries to maximize this objective, and G that tries to minimize it, i.e.:

$$G^* = arg\ min_G\ max_D\ L_{cGAN}\ (G, D). \quad (2)$$

A specific kind of GAN designed for image-to-image translation is the Pix2Pix GAN. Its purpose is to discover a mapping from a source image to an output image, utilizing a dataset that pairs the input and output photos. The key idea behind Pix2Pix is the use of a conditional GAN framework, where both the discriminator and generator networks are depending on the input picture. The discriminator network is a binary classifier that seeks to discern between the created output image and the actual ground truth images from the dataset, whereas the generator networks take the input image as input and attempt to generate the corresponding output image. An adversarial training process is used to develop the discriminator network and generator. The discriminator attempts to accurately identify the generated images and actual ground truth images, while the generator attempts to trick it by producing realistic output images that are identical to matching ground truth photos.

C. Loss Functions

The Pix2Pix GAN, also known as a conditional GAN with Image-to-Image translation, uses a specific objective function that combines adversarial loss with a pixel-wise loss term. Equation (3) is an expression for the pix2Pix GAN's objective function:

$$(G, D) = \lambda_{adv} * L_{adv(G,D)} + \lambda_{pix} * L_{pix} \quad (3)$$

where the generator network is represented by G, and the discriminator network is represented by D, $L_{adv(G,D)}$ is the adversarial network, L_{pix} is the pixel-wise loss,λ_{adv} and λ_{pix} is the hyper parameter that regulates the proportional significance of the pixel-wise and adversarial losses.

Adversarial loss term $L_{adv}(G, D)$ is similar as to standard CGAN objective and motivates the generator to provide realistic results that are indistinguishable from ground truth photos by the discriminator. It can be defined as:

$$L_{adv}(G, D) = E_{x,y}[\log D\ (x, y)] + E_{x,z}[\log(1 - D(x, G\ (x, y))] \quad (4)$$

where x is the ground truth image, z is the noise vector and y is the corresponding output image. The difference between the generated image $G(x)$ and the target image y at the pixel level is measured by the pixel-wise loss term $L_{pix(G)}$

1) Generator

One popular pixel-wise image reconstruction loss that minimizes pixel-level disparities is the L_1loss whose formula is as follows:

$$L_1 = E_{x,y,z}[|y - G(x, y)|] \quad (5)$$

Here, we combined losses MAE loss L_{mae}, Carbonnier loss L_C, SSIM loss L_{ssim} functions i.e.:

$$L_1 = \alpha L_{mae} + \propto L_{carb} + \gamma L_{ssim} \quad (6)$$

By comparing the improved image to the ground truth image, the MAE loss is utilized to assess the quality of the enhanced image. The difference between the desired output and the enhance image can be quantified with the aid of MAE loss. MAE loss is represented by:

$$L_{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (7)$$

The L1 loss and the Carbonnier loss are comparable that they measure the absolute difference between predicted and target values. However, it includes an additional term that introduces a smoothness regularization to the loss function. This regularization term lessens the effect of noise and preserves edges in the rebuilt images. This is how the Carbonnier loss is defined:

$$L_{carb} = \sum x, y, z \sqrt{(y - G(x, y))^2 + \alpha^2} \quad (8)$$

A popular metric for assessing picture quality is called SSIM, which compares the brightness, contrast, and structural similarity between the created and reference images. In this instance, we utilize it to limit the network's training, which is written as:

$$SSIM\_loss = 1 - SSIM(x, y) \quad (9)$$

2) Discriminator

The discriminator loss can be divided into two categories: produced images and real images lost. The discriminator loss can be expressed as:

$$D_{loss} = L_{real} + L_{fake} \quad (10)$$

For real images, the discriminator predicts the labels to be real (1), so the loss is:

$$L_{real} = E[\log D\ (x)] \quad (11)$$

where x is a real image, D represents discriminator network.

For generated output, the discriminator predicts the labels to be fake (0), so the loss is:

$$L_{fake} = E\ [\log(1 - D(G\ (z)))] \quad (12)$$

Here, G represents generator network, z is a latent space input to G.

By combining the two terms, the total discriminator loss is:

$$D_{loss} = E[\log D\ (x)] + E\ [\log(1 - D(G\ (z)))] \quad (13)$$

IV. Experiments

A. Experimental Setup

Dataset and Evaluation: The open-source LOL [10] and SICE [43] datasets are used to assess the effectiveness of the suggested method. The 500 matched low-light and normal-light samples in the LOL dataset were taken especially for studies on low-light image processing. To assess the models, the dataset was divided into 485 training pairs and 15 test pairs. The mapping from low to normal light domains is learned using the training set. SICE dataset comprises two parts, totaling 589 image pairs collected under varied lighting conditions. Part 1 consists of 360 pairs while part 2 contains 229 pairs, each depicting corresponding scenes captured under low and adequate illumination. For model assessment, 560 pairs were allocated to training, whereby the learning of the mapping from low to normal light. The remaining 29 pairs formed the test set, withheld from training for rigorous objective evaluation of generalizability to diverse scenarios not encountered previously.

We made comparisons using state-of-the-art models, such as ZeroDCE [37], MBLLEN [11], DTSD [14], Knowledge Distillation [12], RetinexNet [10], KinD [32], DSENet [44], EnlightenGAN [15], Joint Decomposition and Denoising U-Net (JDD U-Net) [45]. We obtain numerical values by running readily available open-source implementations or extracted data from associated publications. For the numerical evaluation of improved outputs, Structural Similarity Index Measure (SSIM), Learned Perceptual Image Patch Similarity (LPIPS) and Peak Signal-to-Noise Ratio (PSNR) are utilized. Higher SSIM and PSNR indicate better fidelity, while minimized LPIPS indicates higher perceptual accuracy. These objective criteria make it easier to conduct thorough assessments and compare performance to other methods.

Implementation Setup Details: The Keras and TensorFlow framework, which was built in the Python environment and operated on a Kaggle online P100 GPU, was used for all tests. Through the data augmentation procedure, the input photos were arbitrarily cropped, flipped, and formed 400 * 600 patches on both the horizontal and vertical axes. The generator parameters, with a learning rate of 0.1 and a rho of 0.95, were updated using the Adadelta optimizer and the discriminator parameters were updated using the Adam optimizer with a learning rate of 0.0002. The momentum parameters, beta1 and beta2, were adjusted to 0.5 and 0.999, respectively. Under the Pix-to-Pix GAN framework, the model underwent 50 thousand iterations of training. The MAE, SSIM, and Carbonnier loss weights were assigned values of 100, 0.1, and 0.1, in that order.

B. Comparison with State-of-the-Arts

1) Quantitative Evaluation

Using cutting edge techniques, a quantitative assessment of these datasets was first carried out. The numerical outcomes of various techniques are displayed in Tables I and II, where the best results are represented in red, and the second-best results are shown in blue.

As Table I illustrates, on the LOL dataset, our suggested method achieved the maximum SSIM and PSNR score. Furthermore, examining the first and second columns of the table reveals that our model gives the highest values, while the third and fourth columns show the lowest values as compared to others models.

Furthermore, the outcomes in Table II show how much better our suggested method is than other methods on the SICE dataset. Notably, the number of learnable variables and processing requirements of our method are comparatively lower than most competing approaches.

On the whole, our model shows competitive results with less computational resources and then cutting-edge alternative models.

TABLE I
Comparative Analysis of Several Techniques Using the LOL Dataset in Terms of PSNR, SSIM, MAE, and LPIPS Metrics

Model	PSNR↑	SSIM↑	MAE↓	LPIPS↓
KinD	21.1750	0.7684	0.5639	0.4899
ZeroDCE	22.9029	0.7914	0.0398	0.5810
EnlightenGAN	19.4040	0.8212	0.0728	0.4068
RetinexNet	23.0541	0.7706	0.0344	0.4410
MBLLEN	22.2997	0.8989	0.0789	0.4490
DSENet	18.1264	0.4317	0.2586	0.3895
JDD U-Net	18.2169	0.7725	0.8852	0.3587
KD	23.2002	0.8483	0.0409	0.0534
DTSD	22.1672	0.8826	0.0284	0.4490
Ours	25.9442	0.9954	0.0237	0.0459

Note: The top two numerical results are indicated by red and blue highlights, respectively.

TABLE II
Comparative Analysis of Several Techniques Using the SICE Dataset Quantitatively in Terms of PSNR, SSIM, MAE, and LPIPS Metrics

Model	PSNR↑	SSIM↑	MAE↓	LPIPS↓
KinD	19.1750	0.5999	0.0485	0.4970
ZeroDCE	19.0048	0.9320	0.0801	0.5527
EnlightenGAN	18.2709	0.5629	0.0833	0.4273
RetinexNet	22.9851	0.7921	0.0581	0.3907
MBLLEN	22.3828	0.8897	0.0510	0.0988
DSENet	18.8291	0.4995	0.8575	0.3058
JDD U-Net	21.9391	0.7664	0.0628	0.3895
KD	22.5956	0.7866	0.0416	0.4490
DTSD	22.8901	0.7165	0.0316	0.4390
Ours	25.2481	0.9913	0.0282	0.0696

2) Qualitative Evaluation

A qualitative comparison is made the different methods currently in use and the proposed model. Fig. 2 represents the visual outputs obtained by applying various methods using the LOL dataset. As illustrated in Figs. 2 (a) and (b), it is apparent that the methods EnlightenGAN, DTSD and KD exhibit a deficiency in brightness, leading to underexposed images. ZeroDCE, MBLLEN and DSENet tend to over-smooth the image, resulting in a blurring of edge details information. KinD, RetinexNet, JDD U-Net exhibits slight color bias and noise-related issues. In Fig. 2 (c), EnlightenGAN, KinD, RetinexNet and DSENet lead to a loss of edge information details due to excessive blurring. ZeroDCE, MBLLEN and KD exhibit slight color bias and noise information. Figs. 4 (a)-(c) showcase the visual results obtained by applying various methods on the SICE dataset. The results show that the suggested model improves low-light images in an intuitive way and produces visually appealing outputs with proper brightness, consistent colors, and distinct details.

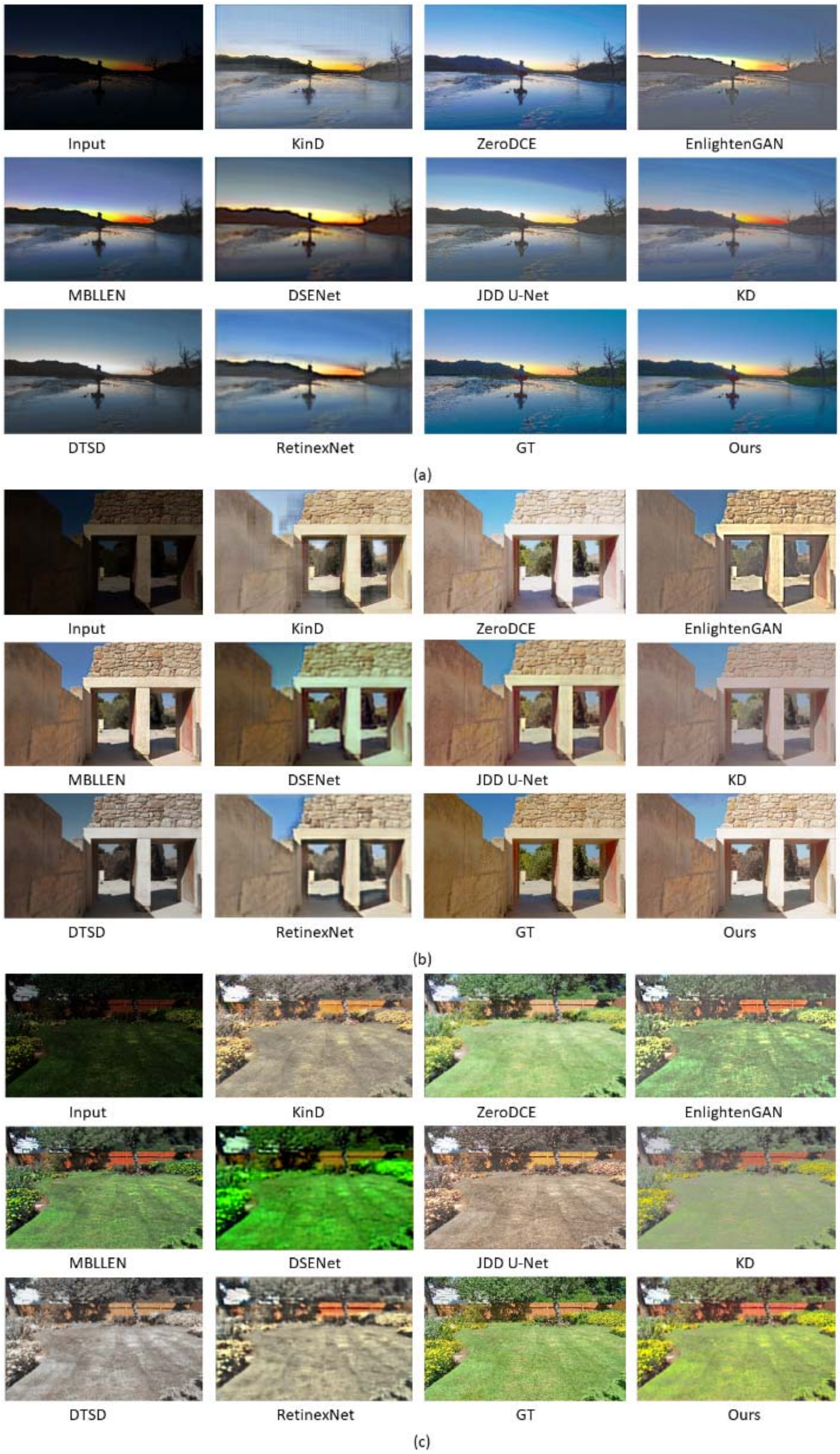

Fig. 4 Qualitative visual comparison outcomes of the various techniques using the SICE dataset

3) Efficiency Evaluation

We compared the computational complexity of the approaches in addition to assessing their performance. This analysis took into account variables including each model's runtime, FLOPs, and number of model parameters. The number of trainable parameters in the model, which measure its spatial complexity, is referred to as the model parameters (in M). On the other hand, FLOPs (in G) serve as a gauge of the model's temporal complexity since they indicate the total amount of floating-point operations that the model has executed. The model's inference time is captured by the runtime (in seconds), and time statistics were performed using the LOL dataset. It provides insights into the efficiency and speed of the model during the inference process. The runtime measurement was conducted on Kaggle with online GPU P100. This information provides context about the system used for running the experiments and obtaining the runtime statistics. The results of the computational complexity are shown in Table III, by utilizing the previous research analysis [46]. This provides a comprehensive overview of the model parameters, FLOPs, and runtime for each method, allowing for a detailed analysis of the computational efficiency of the different approaches. As it can be seen from Table III, our approach has less parameters, FLOPs, and runtime than most other approaches. This demonstrates how the proposed method operates with fewer parameters and allows for faster inference times, all while achieving greater performance. We provide an effective low light picture enhancing approach that balances computational complexity and improved performance.

TABLE III
QUANTITATIVE COMPARISON FINDINGS BETWEEN VARIOUS METHODS' COMPUTATIONAL COMPLEXITY

Model	Flops↓	Runtime↓	Parameter↓
KinD	574.954	0.6725	8.160
ZeroDCE	84.990	0.5758	0.079
EnlightenGAN	273.240	0.7068	8.637
RetinexNet	587.470	0.0740	0.555
MBLLEN	301.120	0.0896	0.950
DSENet	516.971	0.0634	5.923
JDD U-Net	196.359	0. 0618	2.891
KD	83.9530	0.0800	0.283
DTSD	386.640	0.0903	5.270
Ours	79.244	0.0577	0.199

The metrics used for evaluation are runtime (in seconds), FLOPs (in G), and parameters (in M).

V. CONCLUSION

Our study included an alternate autoencoder-based model that we implemented into the Pix-to-Pix GAN framework for low light image improvement. Three autoencoders make up the proposed model, and they cooperate to improve brightness and extract information. The first two autoencoders extract information and brighten the image simultaneously, the third autoencoder is cross-linked with the first two to improve the image that has already been enhanced. By balancing computational complexity and performance, this strategy reduces the burden on performance throughout the testing phase. The advantage of our suggested strategy over other cutting-edge techniques is shown by the outcomes of experiments. Our approach produces better, more realistic images that perform better than those produced by other methods in terms of visual appeal and image quality.

REFERENCES

[1] Badrinarayanan, V., A. Kendall, and R. Cipolla, Segnet: *A deep convolutional encoder-decoder architecture for image segmentation.* IEEE transactions on pattern analysis and machine intelligence, 2017. 39(12): p. 2481-2495.
[2] He, K., et al. Mask r-cnn. in *Proceedings of the IEEE international conference on computer vision.* 2017.
[3] Taigman, Y., et al. *Deepface: Closing the gap to human-level performance in face verification. in Proceedings of the IEEE conference on computer vision and pattern recognition.* 2014.
[4] Schroff, F., D. Kalenichenko, and J. Philbin. *Facenet: A unified embedding for face recognition and clustering. in Proceedings of the IEEE conference on computer vision and pattern recognition.* 2015.
[5] Zuiderveld, K., *VIII. 5. Contrast limited adaptive histogram equalization. Graph.* Gems 474–485 (1994).
[6] Lee, J.-T., et al. *Depth-guided adaptive contrast enhancement using 2D histograms. in 2014 IEEE International Conference on Image Processing (ICIP).* 2014. IEEE.
[7] Land, E.H., *The retinex theory of color vision.* Scientific American, 1977. 237(6): p. 108-129.
[8] Fu, X., et al. *A weighted variational model for simultaneous reflectance and illumination estimation. in Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
[9] Lore, K.G., A. Akintayo, and S. Sarkar, *LLNet: A deep autoencoder approach to natural low-light image enhancement*. Pattern Recognition, 2017. 61: p. 650-662.
[10] Wei, C., et al., *Deep retinex decomposition for low-light enhancement.* arXiv preprint arXiv:1808.04560, 2018.
[11] Lv, F., et al. *MBLLEN: Low-Light Image/Video Enhancement Using CNNs*. in *BMVC*. 2018.
[12] Li, Z., Y. Wang, and J. Zhang, *Low-light image enhancement with knowledge distillation.* Neurocomputing, 2023. 518: p. 332-343.
[13] Fan, S., et al., *LACN: A lightweight attention-guided ConvNeXt network for low-light image enhancement.* Engineering Applications of Artificial Intelligence, 2023. 117: p. 105632.
[14] Zhao, L., et al., *Learning deep texture-structure decomposition for low-light image restoration and enhancement*. Neurocomputing, 2023. 524: p. 126-141.
[15] Jiang, Y., et al., *Enlightengan: Deep light enhancement without paired supervision.* IEEE transactions on image processing, 2021. 30: p. 2340-2349.
[16] Baldi, P., *Autoencoders, Unsupervised Learning, and Deep Architectures, in Proceedings of ICML Workshop on Unsupervised and Transfer Learning,* G. Isabelle, et al., Editors. 2012, PMLR: Proceedings of Machine Learning Research. p. 37--49.
[17] Park, S., et al., *Dual autoencoder network for retinex-based low-light image enhancement.* IEEE Access, 2018. 6: p. 22084-22093.
[18] Goodfellow, I., et al., *Generative adversarial nets.* Advances in neural information processing systems, 2014. 27.
[19] Fu, Y., et al., LE-GAN: *Unsupervised low-light image enhancement network using attention module and identity invariant loss.* Knowledge-Based Systems, 2022. 240: p. 108010.
[20] Stark, J.A., *Adaptive image contrast enhancement using generalizations of histogram equalization.* IEEE Transactions on image processing, 2000. 9(5): p. 889-896.
[21] Lee, C., C. Lee, and C.-S. Kim, *Contrast enhancement based on layered difference representation of 2D histograms.* IEEE transactions on image processing, 2013. 22(12): p. 5372-5384.
[22] Coltuc, D., P. Bolon, and J.-M. Chassery, *Exact histogram specification.* IEEE Transactions on Image processing, 2006. 15(5): p. 1143-1152.
[23] Jobson, D.J., Z.-u. Rahman, and G.A. Woodell, *Properties and performance of a center/surround retinex.* IEEE transactions on image processing, 1997. 6(3): p. 451-462.
[24] Jobson, D.J., Z.-u. Rahman, and G.A. Woodell, *A multiscale retinex for bridging the gap between color images and the human observation of scenes.* IEEE Transactions on Image processing, 1997. 6(7): p. 965-976.
[25] Guo, X., Y. Li, and H. Ling, *LIME: Low-light image enhancement via illumination map estimation.* IEEE Transactions on image processing, 2016. 26(2): p. 982-993.
[26] Guo, X. *LIME: A method for low-light image enhancement. in Proceedings of the 24th ACM international conference on Multimedia.*

2016.

[27] Wang, S., et al., *Naturalness preserved enhancement algorithm for non-uniform illumination images.* IEEE transactions on image processing, 2013. 22(9): p. 3538-3548.

[28] Li, M., et al., *Structure-revealing low-light image enhancement via robust retinex model*. IEEE Transactions on Image Processing, 2018. 27(6): p. 2828-2841.

[29] Fu, X., et al., *A fusion-based enhancing method for weakly illuminated images.* Signal Processing, 2016. 129: p. 82-96.

[30] Wang, W., et al. *Gladnet: Low-light enhancement network with global awareness. in 2018 13th IEEE international conference on automatic face & gesture recognition* (FG 2018). 2018. IEEE.

[31] Wang, R., et al. *Underexposed photo enhancement using deep illumination estimation. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.

[32] Zhang, Y., J. Zhang, and X. Guo. *Kindling the darkness: A practical low-light image enhancer. in Proceedings of the 27th ACM international conference on multimedia.* 2019.

[33] Wang, L.-W., et al., *Lightening network for low-light image enhancement*. IEEE Transactions on Image Processing, 2020. 29: p. 7984-7996.

[34] Zhang, F., et al. *Learning temporal consistency for low light video enhancement from single images. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 2021.

[35] Liu, R., et al. *Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement.* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.*

[36] Jiang, Z., et al., *A switched view of Retinex: Deep self-regularized low-light image enhancement.* Neurocomputing, 2021. 454: p. 361-372.

[37] Guo, C., et al. *Zero-reference deep curve estimation for low-light image enhancement.* in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 2020.

[38] Yang, W., et al. *From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement.* in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 2020.

[39] Yang, Q., et al., *A lowlight image enhancement method learning from both paired and unpaired data by adversarial training.* Neurocomputing, 2021. 433: p. 83-95.

[40] Liu, Y., et al. *PD-GAN: perceptual-details gan for extremely noisy low light image enhancement. in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* 2021. IEEE.

[41] Zhang, J., et al., *DEGAN: Decompose-Enhance-GAN Network for Simultaneous Low-Light Image Lightening and Denoising.* Electronics, 2023. 12(14): p. 3038.

[42] Isola, P., et al. *Image-to-image translation with conditional adversarial networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2017.

[43] Cai, J., S. Gu, and L. Zhang, *Learning a deep single image contrast enhancer from multi-exposure images.* IEEE Transactions on Image Processing, 2018. 27(4): p. 2049-2062.

[44] Singh, K. and A.S. Parihar, *DSE-Net: Deep simultaneous estimation network for low-light image enhancement*. Journal of Visual Communication and Image Representation, 2023. 91: p. 103780.

[45] Deng, J., et al. *Low-light Image Enhancement based on Joint Decomposition and Denoising U-Net Network.* in *2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*. 2020. IEEE.

[46] C. Li et al., "*Low-Light Image and Video Enhancement Using Deep Learning:* A Survey," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 9396-9416, 1 Dec. 2022, doi: 10.1109/TPAMI.2021.3126387.