Brightness-Aware Generative Adversarial Network for Low-Light Image Enhancement

Huafei, Zhao, Mideth Abisado
College of Computing & Information Technologies, National University, Manila, Philippines

***Abstract*—Images from low-light frequently exhibit poor visibility, excessive noise, and color distortion, which substantially impair both computer vision systems and human visual perception. Although numerous enhancement techniques have been developed, producing visually appealing results with well-maintained structural details and natural color reproduction continues to pose significant challenges. To address these limitations, this paper present an Brightness-Aware Generative Adversarial Network (BA-GAN) for robust low-light image enhancement (LLIE). Our framework employs a U-Net-based generator that effectively captures multi-scale contextual features while preserving fine image details through skip connections. The key innovation lies in our novel Brightness Attention Mechanism Module, integrated within the decoder, which dynamically directs the network's focus to regions requiring substantial illumination correction. To ensure local photorealism, this paper adopt a PatchGAN discriminator architecture. The complete model is trained on the LOL dataset using a composite loss function combining: (1) adversarial loss for realistic image generation, (2) brightness attention loss for keeping the brightness accuracy, and (3) perceptual loss to maintain structural and semantic fidelity. Extensive experiments validate that our BA-GAN outperforms current state-of-the-art methods, achieving superior performance on both quantitative metrics (PSNR: 20.7127, SSIM: 0.7963, LPIPS: 0.2271) and qualitative visual assessments. The enhanced images demonstrate significantly improved visibility while effectively suppressing noise and preserving natural color characteristics.**

Keywords*—*Low-light image enhancement; generative adversarial networks; U-Net; PatchGAN; attention mechanism; deep learning

I. Introduction

Images captured by cameras often used by advanced computer vision tasks such as image classification, object detection, and image generation, and security monitoring [1]. However, during the capture process, due to low-light environments or insufficient device lighting capabilities, the resulting images are often dim, visually unclear, and have low contrast. These images significantly hinder subsequent computer vision processing, rendering them largely unusable. Low-quality images not only affect human visual experience but also have adverse effects on subsequent computer vision applications.

When capturing images in low-light environments, long exposure or slow shutter speeds are often used to increase the amount of light captured by the camera, making more complex low-light blur degradation unavoidable [2]. While flash photography is commonly employed to brighten low-light environments, it often introduces uneven illumination artifacts that degrade image quality through increased noise and unnatural lighting patterns. To address these issues, professional photographers typically enhance the quality of pictures taken under low-light environments by adjusting camera parameters or using image editing software (such as Adobe Photoshop) [3]. However, this process usually requires complex operations and professional skills, making it unsuitable for ordinary users. Therefore, automated LLIE technology is crucial for solving this problem.

Conventional approaches to LLIE mainly consist of histogram equalization techniques [4] and Retinex-based methods [5]. With the advancement of GPU computing power, the widespread application of neural networks and deep learning, and the emergence of training datasets, it is now possible to learn detailed information from images through deep learning methods, thereby improving the contrast of images with low-light [6]. Deep learning methods utilize deep neural networks to establish an end-to-end mapping between low-light and normal-light image domains through data-driven learning from extensive training samples. Compared to traditional methods, they possess more powerful feature representation capabilities and have achieved impressive results on benchmark datasets [7]. Existing LLIE methods, while capable of upgrading the brightness and contrast of dim images to some extent, still have issues such as color distortion, noise amplification, and insufficient model generalization [8] as it's hard to obtain paired datasets in real-world production and life, as well as the limited global processing capability of convolutional operations on images.

GANs [9] have proven effective for image-to-image translation tasks, including low-light enhancement. GANs leverage a generator network to produce enhanced images and a discriminator network to distinguish generated fake images from real normal-light images, leading to perceptually more convincing results. However, existing methods may still struggle with balancing noise suppression, detail preservation, and adaptive enhancement across different image regions and lighting variations.

To solve these limitations, this paper propose a Brightness-Aware Generative Adversarial Network (BA-GAN). The key contributions in this paper summarized as follows:

1) A U-Net generator architecture that fuses multi-scale details. This paper adopted a U-Net-based generator, which processes images through a symmetric encoder-decoder structure. The key to this architecture lies in its skip connections. These connections directly transmit the low-level, high-

*Corresponding Author

resolution detailed features (such as edges and textures) captured by the encoder at different levels to the corresponding levels of the decoder. This design can effectively avoid the common problem of detail information loss in deep networks, thereby maximizing the preservation of fine structures while enhancing the image brightness and preventing the final result from becoming blurred.

2) A novel illumination attention module for adaptive enhancement. This is the core innovation of this paper. We integrated an illumination attention module into the decoder of the generator. This module aims to solve the problem of uneven illumination in low-light images, enabling the network to dynamically and adaptively focus its attention on the area's most in need of illumination correction according to the image content. By learning and applying this spatially varying enhancement strategy, the model can avoid overexposing areas with existing light sources while sufficiently brightening extremely dark areas, thus achieving a more natural and accurate enhancement effect.

3) A PatchGAN discriminator that focuses on local realism. To generate more perceptually convincing results, our framework employs a PatchGAN discriminator. Different from traditional discriminators that make true/false judgments on the entire image, PatchGAN evaluates the authenticity by assessing local patches of the image. This mechanism encourages the generator to not only focus on the global visual effect but also pay attention to generating locally realistic textures and structures. This helps to significantly improve the detail quality and overall realism of the enhanced image.

4) A composite loss function for balanced optimization. The low-light enhancement task requires a balance among multiple objectives. To this end, this paper designed and adopted a composite loss function to guide the training of the model. This function combines three key components: (1) adversarial loss, used to enhance the realism of the generated image; (2) brightness attention loss, used to ensure the accuracy of brightness correction; and (3) perceptual loss, used to maintain the original structure and semantic content of the image in the deep feature space. This multi-objective optimization strategy ensures that the final enhancement result achieves the best balance in terms of visual realism, brightness, and content fidelity.

In summary, LLIE is crucial for improving visual quality and enabling robust performance in downstream computer vision tasks. However, existing methods ranging from traditional techniques like histogram equalization and Retinex theory to deep learning approaches—often suffer from color distortion, noise amplification, and limited generalization due to the scarcity of real-world paired datasets and the inherent constraints of convolutional operations. While GANs have shown promise in LLIE, challenges remain in achieving adaptive enhancement, detail preservation, and noise suppression. To address these limitations, this paper propose a BA-GAN that integrates a U-Net generator with multi-scale skip connections, a novel illumination attention module for adaptive brightness correction, a PatchGAN discriminator for local realism, and a composite loss function for balanced optimization. Our approach aims to deliver natural, high-fidelity enhancements while addressing uneven illumination and preserving fine-grained details.

II. Related Works

A. Traditional Methods

Conventional approaches to LLIE mainly consist of histogram equalization techniques and Retinex-based methods.

The histogram equalization method operates on the histogram of an image and upgrades the image's contrast by altering its dynamic range, which can enhance the quality of dim images to some extent [10]. The histogram records the distribution of image data, and due to its uneven distribution, this method redistributes it to achieve non-linear stretching, making the distribution of image data more reasonable. Examples include HE [2] and CLAHE [11]. Numerous improved methods based on HE, have enhanced HE to varying degrees. Compared to global equalization, local equalization focuses more on the internal detail distribution of the image, but the required computational effort also increases. Low-light images enhanced by such methods are prone to issues like overexposure or underexposure, color distortion, artificial artifacts, loss of image details, and unnatural appearance. Sometimes, the resulting images may also retain the noise present in the original images.

Retinex-based methods perform image enhancement through a decomposition process, separating low-light images into illumination and reflectance components. This separation is typically achieved by applying physical priors or mathematical regularization constraints. The reflectance component is subsequently obtained by solving an optimization problem involving both the estimated illumination and the original input image [12], and is considered the final enhancement result. Building on Retinex, Single Scale Retinex (SSR) [13] was proposed, which removes the influence of illumination by comparing the central element with the surrounding area's luminance values. Multi-Scale Retinex (MSR) [14] was introduced to address the issue of Gaussian kernel optimization quality. The MSR with Color Restoration (MSRCR) [15] algorithm adjusts color distortion by adding a color restoration factor.

Retinex-based LLIE methods can effectively address issues such as underexposure and overexposure. However, they often overlook the presence of noise during image processing, leading to noise amplification. Additionally, these methods struggle to find an effective prior or regularization. As a result, low-light images enhanced by these methods often have limitations like color distortion, artificial artifacts, overexposure, and accompanying noise.

B. Deep Learning-Based Methods

The rapid evolution of deep learning has led to the development of numerous data-driven approaches for LLIE, consistently outperforming conventional methods in terms of image quality improvement. These techniques leverage deep neural architectures to learn complex mappings between low-light and normal-light image domains, effectively enhancing visual clarity and perceptual quality. Current deep learning-based approaches can be systematically categorized into four paradigms: (1) supervised learning, (2) unsupervised learning, (3) semi-supervised learning, and (4) zero-shot learning methods.

1) Supervised learning methods. Supervised learning is widely used in LLIE. It involves providing a set of images along with their corresponding enhanced versions as training data, enabling the model to learn an end-to-end transformation that converts input images into enhanced outputs with improved visual quality. As the pioneering deep learning solution for LLIE, LLNet [16] demonstrated that a carefully designed autoencoder could effectively enhance low-light images while preserving important visual features. MSR-NET [17] introduced the use of convolutional neural networks, combined with Retinex theory, to restore images to good brightness levels. Wei et al. [18] proposed Retinex-Net, which includes decomposition, adjustment, and reconstruction, corresponding to three sub-networks. During the adjustment phase, the brightness of the illumination map is increased through multi-scale encoders and decoders. In the final reconstruction stage, the enhanced image is obtained through element-wise multiplication of the processed illumination and reflectance maps. However, the edges of the enhanced structure are overly prominent, and the naturalness of the entire image is not perfect. KinD [19] followed the Retinex theory-based approach of decomposition followed by enhancement, constructing a two-stage network. The restoration sub-network introduced information from real illumination to prevent image distortion, and the loss function ensured the smoothness of the illumination layer. KinD++ [20] alleviated residual visual artifacts in the results generated by KinD through a multi-scale illumination attention module. MIRNet [21] used a multi-scale feature fusion method, including multi-scale residual blocks, multi-resolution convolutional streams, and attention mechanism-based multi-scale features, achieving high performance in image enhancement tasks. Liu et al. [22] proposed a three-stage brightness-aware network based on brightness-aware attention and residual quantization encoding blocks, and their designed a query module link the low-light and normal-light domains. Building upon U-Net [23], LAU-Net [24] augments the baseline architecture with a Parallel Attention Unit (PAU) to weight informative features dynamically, an Internal Resizing Module (IRM) for scale-invariant representation learning, and auxiliary convolutional layers to mitigate low-light artifacts.

Supervised learning methods have shown improvements in color balance and brightness adjustment compared to traditional methods, but they tend to lose local details and often rely on paired datasets, leading to overfitting and low cross-data generalization. In the field of LLIE, paired datasets are limit in quantity and synthesizing image datasets can result in poor model generalization.

2) Unsupervised learning methods. Unsupervised learning refers to training data without corresponding target enhanced images, allowing the model to learn the structure and features of image data on its own. After the introduction of GANs, GANs achieved significant results in computer vision tasks. This method do not need paired training data and can be trained using unpaired datasets, making data acquisition relatively easier. In the realm of LLIE, EnlightenGAN proposed by Jiang et al. [25] was the first to eliminate the dependency on paired training data, greatly enhancing the flexibility of generated images and adapting to various scenarios. The introduction of RetinexGAN [26] combined GANs with the Retinex model, incorporating a decomposition network and two discriminative networks, where the decomposition network decomposes the Retinex model, and the discriminators evaluate the illumination and reflectance components. Zhu et al. [27] designed Cycle-consistent Generative Adversarial Networks (CycleGAN), which use cycle-consistency loss to achieve image style transfer. However, applying it to LLIE training is challenging as it overlooks the preservation of local features. Hu et al. [28] developed a novel two-stage unsupervised framework that separates the enhancement process into distinct pre-enhancement and post-enhancement phases. The initial stage employs conventional Retinex-based algorithms for primary image enhancement, while the subsequent stage utilizes an adversarial trained refinement network to achieve superior quality improvement. In a related approach, Wang et al. [29] introduced an attention-guided unsupervised GAN architecture featuring: (1) an auxiliary edge restoration module for enhanced sharpness preservation, and (2) a dedicated attention mechanism for improved colour fidelity reconstruction.

3) Semi-supervised learning methods. Semi-supervised learning integrates the advantages of supervised and unsupervised learning. Yang et al. [25] constructed a deep recursive frequency band network, which first uses supervised learning to learn frequency band representations from coarse to fine, and then employs unsupervised adversarial learning to improve the model's generalization ability. Through this end-to-end recursive training, noise removal is effectively achieved, and the ability to recover local structural details is improved. Dimma [30] built a brightness adjustment module to obtain a low-light version of normal-light images, thereby enabling semi-supervised training.

4) Zero-shot learning methods. Zero-shot learning in the field of LLIE refers to the ability to learn how to enhance images solely from test images. Guo et al. introduced Zero-DCE [31], a lightweight, reference-free deep curve estimation model that achieves pixel-level adjustments of input images through image-to-curve mapping and utilizes a loss function to drive reference-free training. Building on this, an accelerated lightweight version, Zero-DCE++ [32], was introduced. Zhang L et al. introduced a "zero-learning" approach for backlit image restoration, which is based on deep learning but does not rely on any training image. The framework ExCNet, a specially designed convolutional neural network architecture. Zhu et al. [32] introduced RRDNet, an innovative three-branch CNN that decomposes input images into illumination, reflectance, and noise components. This architecture employs iterative optimization of a custom loss function to simultaneously predict noise patterns and restore underexposed regions. In a complementary approach, Liu et al.'s RUAS [33] formulates a mathematical model to represent the intrinsic illumination structure of low-light images, then unfolds this optimization process into a learnable network architecture. By exploring optimal configurations within a constrained search space, RUAS achieves superior enhancement performance through its propagated structure.

In summary, conventional enhancement methods depend on fixed statistical priors, limiting their adaptability to diverse real-world scenarios. While contemporary supervised approaches predominantly employ U-Net architectures, they frequently suffer from spatial detail degradation and inadequate feature representation due to the absence of attention mechanisms. To address these limitations, our proposed framework innovatively combines attention modules with a U-Net backbone and incorporates a PatchGAN discriminator. This hybrid architecture enables: (1) enhanced multi-scale feature learning, (2) improved discriminative feature extraction, (3) effective noise suppression, and (4) robust contextual information aggregation across different scales.

III. Methodology

The LLIE method based on the GANs model introduced in this paper employs U-Net as the generator, which is characterized by its low data requirement and fast training speed. By using PatchGAN as the discriminator instead of the traditional binary classifier, the trained model pays more attention to image details, improves image quality, and accelerates network convergence. The network structure of the method used is illustrated in Fig. 1. The generator consists of a U-Net backbone with embedded illumination attention modules. The discriminator follows a PatchGAN structure to promote local realism.

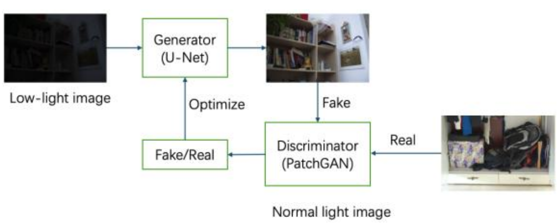

Fig. 1. The proposed BA-GAN structure.

C. Generator

U-Net has demonstrated success in image segmentation, image translation, and enhancement. Through the structure of the encoder and decoder, along with skip connections, U-Net avoids generating blurry images or losing details (such as edges and textures). The symmetric structure of U-Net better preserves spatial correspondences.

The extracted image features are encoded and decoded by the U-Net network, and the same convolution and structural detail residual fusion operations are performed on all branches to improve the features, resulting in feature maps with stronger expressive capabilities. The U-Net network can combine low-level and high level features from the encoder and decoder, fully utilizing contextual texture information. The proposed generator structure is illustrated in Fig. 2.

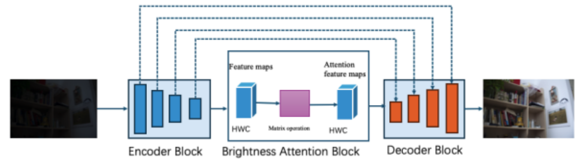

Fig. 2. The structure of the proposed generator.

D. Brightness Attention Module

In the low-light enhancement task, the degradation of the image is not uniformly distributed. Some areas may be almost completely black and require significant brightening, while other areas may have a small amount of light sources and need fine adjustment to avoid overexposure. Due to its fixed local receptive field, the traditional convolution operation has difficulty capturing such global and dynamic illumination dependencies. Therefore, this paper designed an Brightness Attention Module (BAM). Its core objective is to enable the network to adaptively evaluate the importance of different feature channels for the final illumination restoration and recalibrate the features accordingly, so as to achieve precise and differentiated enhancement of illumination.

Our BAM is strategically placed between the encoder and decoder of the U-Net. It takes the deepest feature map $F_{enc} \in R^{C\times H'\times W'}$ output by the encoder as input and performs the following three consecutive operations to generate the attention-calibrated feature map $F_{att} \in R^{C\times H'\times W'}$.

1) Illumination squeeze. First, this paper perform Global Average Pooling on the input feature map F_{enc}, compressing it in the spatial dimension to generate a scalar descriptor for each channel. This process can be regarded as the extraction and generalization of the global illumination information contained in each feature channel. The object is to obtain a vector $z \in R^C$, where each element z_c represents the global illumination response of the c-th feature channel.

$$\mathrm{z_c} = \mathrm{F_{sq}}(\mathrm{F_{enc}}) = \frac{1}{\mathrm{H'\times W'}}\sum_{\mathrm{i=1}}^{\mathrm{H'}}\sum_{\mathrm{j=1}}^{\mathrm{W'}}\mathrm{F_{enc_c}}(\mathrm{i,j}) \quad (1)$$

2) Illumination excitation. To learn the complex non-linear illumination dependencies between channels, this paper uses two fully connected (FC) layers to process the compressed descriptor z. The first FC layer reduces the dimension from C to C/r (where r is the reduction ratio) and uses the ReLU activation function. The second FC layer restores the dimension to C and uses the Sigmoid activation function, ultimately generating the attention weights $s \in R^C$ for each channel. The output values of the Sigmoid function range from 0 to 1, which can exactly serve as the "importance" or "gain" weights for each channel. The closer the value is to 1, the more important the channel is for illumination restoration.

$$s = F_{ex}(z, W) = \sigma\big(W_2\delta(W_1 z)\big) \quad (2)$$

Where σ is the Sigmoid function, δ is the ReLU function, and W_1 and W_2 are the weights of two fully connected layers.

3) Feature recalibration. Finally, this paper multiplies the learned channel attention weights s with the original input feature map F_{enc} channel by channel to obtain the final output attention feature map F_{att}. This operation enables the network to enhance the feature channels that are beneficial for illumination enhancement while suppressing the irrelevant channels that may introduce noise or artifacts. This calibrated feature map F_{att} is then fed into the decoder to guide the subsequent image reconstruction process.

$$F_{att_c} = s_c \cdot F_{enc_c} \tag{3}$$

Although the BAM this paper proposed draws on the idea of channel attention in structure [34], its uniqueness and effectiveness are reflected in the following aspects:

- Task Specificity: Different from general attention modules, our BAM is specifically designed to solve the problem of uneven illumination. By modeling global information at the bottleneck of the network, it focuses on learning which regional features should be brightened and which dark - area details should be retained, rather than general feature enhancement.
- Strategic Placement: This paper place the BAM at the bottleneck of the U-Net, instead of spreading it after each convolutional block as in traditional practices. The reason for this design is that performing global illumination calibration at the place where information compression is most extreme allows the decoder to obtain prior knowledge about the overall illumination distribution at the first step of the reconstruction process, thereby more effectively guiding all subsequent upsampling and feature fusion steps and fundamentally reducing artifacts and color distortion.
- Synergy with Loss Function: More importantly, the design of this module forms a synergy with the Brightness Attention Loss function this paper proposed. The BAM learns the intrinsic dependence of illumination at the feature level, while the brightness loss supervises the final brightness accuracy at the pixel level. The two jointly guide the network from different dimensions, ensuring the realism and accuracy of illumination enhancement.

E. Discriminator

In designing the discriminator network, this paper references the PatchGAN discriminator proposed by Isola et al. [27] to replace the traditional binary classifier discriminator. This discriminator outputs an N × N matrix, and the image is judged as real or generated based on the mean value of the matrix. Compared to the traditional binary classifier discriminator, the advantage of using the PatchGAN discriminator lies in its output being a matrix, with the final result being the matrix mean, which fully considers the influence of different regions of the image, thereby enhancing image quality and paying more attention to image details. Additionally, the computation of small-sized image patches significantly accelerates the convergence speed of the network. This paper designs a CNN model with 4 fully convolutional layers. Except for the last convolutional layer, the other 3 layers undergo BatchNorm for data standardization after convolution, and LeakReLu is used as the activation function. The proposed discriminator structure is illustrated in Fig. 3.

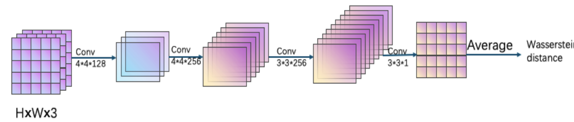

Fig. 3. The structure of the discriminator.

F. Loss Functions

Adversarial Loss [35]. The adversarial loss is the core driving force of the GANs framework. Its fundamental goal is to drive the generator to produce enhanced images that are more visually realistic and indistinguishable from real normal - illuminated images in terms of distribution. This process can be understood as a two - player game involving two networks (the generator G and the discriminator D).

Generator (G): The generator is a U-Net-based network. Its task is to receive a low-light image as input and output an enhanced fake image I_{gen}. The goal of the generator is to deceive the discriminator as much as possible and make it believe that the generated image is real.

Discriminator (D): The discriminator adopts the PatchGAN architecture. Its role is like an image connoisseur or a binary classifier, and its sole goal is to accurately distinguish between the real normal-illumination image I_{real} and the fake image I_{gen} produced by the generator.

$$L_{adv} = E\left[log\left(D\left(I_{real}\right)\right)\right] + E\left[log\left(1 - D\left(I_{gen}\right)\right)\right] \tag{4}$$

Compared with traditional loss functions that only rely on pixel-level differences (such as L1 or MSE loss), adversarial loss has significant advantages in improving the perceptual quality of images. It does not require the generated image to be exactly the same as the target image in terms of pixels, but encourages the generated image to look more realistic in terms of texture, structure, and overall style.

In particular, the PatchGAN discriminator used in this paper does not output a single true/false judgment for the entire image. Instead, it makes judgments on N×N-sized image patches (Patches) in the image and then averages all the judgment results. This mechanism forces the generator to not only focus on global consistency but also ensure that the local texture and structure of the image are equally realistic, thereby effectively improving the quality of image details and accelerating the convergence of the network.

Brightness Attention Loss. Using global L1 or L2 loss directly between the brightness and the target image treats all regions in the image equally. However, in the low-light enhancement task, the human eye perceives brightness errors differently in different regions [36]. For example, in regions that already have a light source or are rich in texture, slight deviations in brightness are more likely to be noticed and cause artifacts; while in smooth and extremely dark regions, the weight of the error can be relatively lower [37]. To make the model prioritize the key regions that have a greater impact on visual quality during the optimization process, this paper designed the Brightness Attention Loss Function.

Attention Map Generation. This paper assumes that in regions with drastic luminance changes (i.e., large gradient values), such as edges and textures, the accuracy of luminance restoration is crucial for the overall perceptual quality [38]. Therefore, this paper uses the gradient magnitude of the luminance channel of the target image to generate the attention map A. The specific steps are as follows:

First, this paper uses the standard Sobel operator to calculate the gradients G_x and G_y of the target luminance map Y_{target} in the horizontal and vertical directions respectively.

Then, this paper calculates the gradient magnitude of each pixel.

$$M_{grad}(i,j) = \sqrt{G_x(i,j)^2 + G_y(i,j)^2} \tag{5}$$

To convert the gradient magnitude into a normalized attention weight map A this paper processes it to make its values more concentrated in the key regions. First, this paper normalizes it to the range of $[0,1]$, then add a small constant ε to avoid division by zero, and use it as the final attention weight. The weight map A generated in this way has higher values in the edge and detail regions of the image and lower values in the smooth regions.

$$A(i,j) = \frac{M_{grad}(i,j)}{\max(M_{grad})} + \varepsilon \tag{6}$$

Finally, this paper combines the above attention weight map A with the brightness difference between the generated image Y_{gen} and the target image Y_{target} to define our brightness attention loss.

$$L_{bright} = \frac{1}{H \times W} \sum_{i,j} A(i,j) \cdot \left(Y_{gen}(i,j) - Y_{target}(i,j)\right)^2 \tag{7}$$

This paper uses the weighted L2 loss (squared difference) here mainly for two reasons: First, compared with the L1 loss, the L2 loss imposes a stronger penalty on larger errors, which helps the model quickly correct obvious brightness deviations in the early stage of training. Second, by multiplying with the attention weight A, this loss function forces the generator to prioritize optimizing the brightness accuracy in the key regions of the image structure (such as edges and textures), thereby effectively preserving image details and suppressing artifacts while improving the overall brightness.

Perceptual Loss [35] mainly uses the differences between high-dimensional feature maps of images to guide model training. Perception loss can make the image more similar in high-level information such as content and global structure, ensure the semantic information of the image as much as possible, and also enhance the detail information, which is conducive to the training of unpaired data sets.

$$L_{percep} = \frac{1}{W_i H_i C_i} \| \varphi_i(I_x) - \varphi_i(G(I_x)) \| \tag{8}$$

Among them, the φ_i represents the ith feature map taken from VGG16 network, I_x represents the input low illumination image, in the feature map, W is the width, H is the height and C is the channel.

Generator loss [9]. In the BA-GAN framework proposed in this paper, the training objectives of the generator are multi-faceted. It not only needs to enhance the image brightness but also ensure the realism of the results and the fidelity of the content. To achieve this complex balance, the optimization of the generator is not driven by a single objective but is guided by a composite loss function. This total loss function L_G is a weighted sum of three independent loss terms, and each term targets a specific aspect of the enhancement process.

$$L_G = \lambda_{adv} * L_{adv} + \lambda_{bright} * L_{bright} + \lambda_{percep} * L_{percep} \tag{9}$$

Among them, $\lambda_{adv}, \lambda_{bright}, \lambda_{percep}$ are weights within the composite generator loss function. Through the synergistic effect of these three loss functions, the generator learns how to strike a balance among multiple objectives during the training process: making the image brighter and accurate, making it look real and natural, and at the same time retaining all the content of the original scene. The weights for the composite loss function were empirically set to λ_{adv}=1, λ_{bright}=1, and λ_{percep}=0.1 to balance the contributions of each component.

Discriminator loss [9]. After each iteration, calculate the loss function to get the difference between the fake and true images, then guide the next step of training. The specific method is to first initialize the model parameters with random values, input a low light image, obtain an enhanced image through a generation network, calculate the error between images using a discriminator, pass the error back along the direction of minimum gradient, modify the parameter values, repeat multiple times until the error value reaches a satisfactory value, and then stop iterating to obtain the final required model.

In the GANs framework proposed in this paper, the discriminator plays the role of an image connoisseur. The discriminator loss (L_D) is the core indicator for measuring the level of its discrimination ability. Its fundamental goal is to drive the discriminator network to more and more accurately distinguish between real normal-illuminated images and enhanced images forged by the generator.

The training process of the discriminator receives two types of inputs:

Real images I_{real}: Real, high-quality normal-illumination images from the LOL dataset.

Fake images I_{gen}: Enhanced images output by the U-Net generator after processing low-light images.

The goal of the discriminator is to output a probability value close to 1 for real images (judged as true) and a probability value close to 0 for fake images (judged as false). To achieve the above goals, the training of the discriminator aims to maximize its loss function L_D, which is defined in the paper as follows:

$$L_D = E[\log D\,(I_{real})] + E\left[\log\left(1 - D\left(I_{gen}\right)\right)\right] \tag{10}$$

In each training iteration, the model calculates the value of this loss function to measure the difference between the real image and the fake image. Then, this error is backpropagated along the gradient to update the network parameters of the discriminator itself.

Different from traditional discriminators that output a single true/false probability for the entire image, PatchGAN divides the input image into multiple N×N image patches (Patches) and makes a authenticity judgment for each patch, ultimately outputting an N×N matrix. The final judgment result is obtained by taking the average of this matrix. The advantage of this mechanism is that it forces the discriminator to focus on the local details and texture realism of the image, rather than just the global appearance, which helps to improve the overall quality of the generated images and accelerate the convergence of the network.

IV. Experiments

A. Dataset

The existing LOL-dataset [39] is used in this study. The LOL-dataset includes 485:15 pairs of images used for training and testing. The most core feature of the LOL-dataset is that it contains paired images. Each pair consists of an image taken under low-light conditions and its corresponding ground-truth image of the same scene taken under normal lighting conditions. This paired characteristic makes it very suitable for calculating the metrics of model performance. The test set of this dataset is used as a standard platform to objectively evaluate and compare the performance of the method in this paper with other representative algorithms. Below image pairs in Fig. 4 are some of the image samples from the LOL-dataset.

Fig. 4. The image samples from the LOL-dataset.

B. Experimental Setting

Hardware and platform: All training and testing were completed on a server equipped with an NVIDIA GeForce 4090 GPU, and the operating system was Ubuntu.

Software framework: Our model was implemented based on the PyTorch deep learning framework.

Training hyperparameters: The model was trained using the Adam optimizer. The initial learning rate was set to 1e-4, the batch size was set to 64, and the model was trained for a total of 3000 epochs.

Algorithm 1: Training of the proposed BA-GAN Algorithm

Input: Learning rate α, Gradient penalty coefficient λ; Adam optimizer parameters α, β_1, β_2; batch size m; low-light image x_i;

Initialize: discriminator parameters w_0; generator parameters θ_0

While (generator parameter θ not converged do) **do**

For (each epoch) **do**

Sample a batch of m low-light images $\{x^{(i)}\}_{i=1}^{m}$ from the dataset.

Sample a batch of m real normal-light images $\{y^{(i)}\}_{i=1}^{m}$ from the dataset.

Generate fake images: $\tilde{x}^{(i)} = G_\theta(x^{(i)})$.

Calculate the discriminator loss L_D:

$$L_D \leftarrow E[\log D_w(y^{(i)})] + E[\log(1 - D_w(\tilde{x}^{(i)}))]$$

Update discriminator parameters via gradient ascent:

$$w \leftarrow w + \alpha \cdot \text{Adam}(\nabla_w L_D, w, \beta_1, \beta_2)$$

End for

Sample a batch of m low-light images $\{x^{(i)}\}_{i=1}^{m}$ from the dataset.

Calculate the composite generator loss L_G:

$$L_G \leftarrow \lambda_{adv} L_{adv} + \lambda_{bright} L_{bright} + \lambda_{percep} L_{percep}$$

Update generator parameters via gradient descent:

$$\theta \leftarrow \theta - \alpha \cdot \text{Adam}(\nabla_\theta L_G, \theta, \beta_1, \beta_2)$$

End **Output**: The trained generator G_θ.

C. Evaluation Metrics

To more objectively assess the performance of the improved model in this paper, common image assessment algorithms were used to evaluate some representative methods. The evaluation metrics selected were Peak Signal-to-Noise Ratio (PSNR) [40], Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) [41].

PSNR is a widely used metric for measuring image quality. It evaluates image quality based on the errors between corresponding pixels. This metric quantifies the degree of image distortion by calculating the ratio of the signal (the maximum possible pixel value of the image) to the noise (the mean squared error between the enhanced image and the real image). When evaluating the results, a higher PSNR value indicates better quality and less distortion of the enhanced image.

The calculation of PSNR is based on the Mean Square Error (MSE) [42]. MSE is calculated first, and then PSNR is calculated according to MSE. The calculation formula of MSE is:

$$MSE = \frac{1}{\text{nm}} \sum_{i=1}^{\text{m}} \sum_{j=1}^{\text{n}} \left(I_{(i,j)} - K_{(i,j)}\right)^2 \qquad (11)$$

Where m and n are the number of rows and columns of the image, $I(i,j)$is position that the pixel value of the original image, $K(i,j)$ is the pixel value of the processed image at the same position. After obtaining MSE, PSNR can be calculated by the following formula:

$$PSNR = 10 \log_{10}\left(\frac{\text{MAX}_\text{I}^2}{\text{MSE}}\right) \qquad (12)$$

Among them, MAX I is the maximum possible image pixel value. For 8-bit images, MAX I is usually 255.

The SSIM quantifies image similarity through three key components: luminance, contrast, and structural composition. Higher SSIM values (ranging from 0 to 1) indicate greater similarity between the enhanced and reference images.

$$SSIM(x,y) = \left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}\right) \cdot \left(\frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}\right) \qquad (13)$$

X and y are the two images to be compared, μ_x And μ_y Average brightness of image x and y, $\mu_x{}^2$ And $\mu_y{}^2$ Is the variance of image x and y, μ_{x_y} is the covariance of image xx and y, and C1 and C2 are small constants added to avoid zero denominator.

LPIPS is a modern metric that measures perceptual similarity by simulating human visual evaluation. It uses a deep learning network to extract and compare the deep features of image patches, thereby determining the similarity between two images in the eyes of human observers. Different from the previous two metrics, a lower LPIPS value (closer to 0) represents higher perceptual fidelity between the two images, meaning they are more similar visually.

V. Results and Discussions

A. *Comparsion with other Methods*

To evaluate the effectiveness of our method, this paper compares it with several representative low-light enhancement algorithms on the LOL dataset, including HE [11], MSR [14], Retinex-Net [18], EnlightenGAN [25], KinD [19], Zero-DCE [31], and SCI [43]. Quantitative comparisons of these metrics across different methods are presented in Table I.

As shown in Table I, our proposed method achieves the highest SSIM score (0.7963) and the best PSNR (20.7127 dB) among all compared methods, indicating superior structural preservation and noise suppression capabilities. Moreover, our method attains a competitive LPIPS score (0.2271), demonstrating enhanced perceptual quality compared to most baseline approaches. These results validate the effectiveness of integrating U-Net, PatchGAN, and illumination attention for robust LLIE.

Fig. 5 presents qualitative comparisons of enhanced results on representative samples from the LOL dataset using different methods. Compared to EnlightenGAN, Zero-DCE, and SCI, the proposed method achieves superior visual quality in terms of brightness balance, detail preservation, and color fidelity.

Specifically, EnlightenGAN tends to produce over-saturated regions and amplifies noise in extremely dark areas. Zero-DCE improves brightness but suffers from local color distortion and residual noise, particularly in high-texture regions (e.g., books, patterns). SCI shows smoother brightness but lacks sharpness in edges and tends to flatten subtle textures.

In contrast, the proposed method generates enhanced images that are perceptually more natural and visually closer to the ground truth. The edges and structural details (e.g., in cabinets, text, and metal objects) are well preserved, and the illumination is more uniformly distributed across the image. Furthermore, color consistency is notably better, with reduced color cast and halo artifacts compared to the other approaches.

TABLE I EVALUATION RESULTS ON LOL DATASET

Method	SSIM↑	PSNR↑	LPIPS↓
HE [11]	0.6433	19.9959	0.3551
MSR [14]	0.5327	11.4331	0.3505
Retinex-Net [18]	0.7903	18.9775	**0.2036**
EnlightenGAN [25]	0.7117	15.5693	0.2547
KinD [19]	0.6475	16.0338	0.2979
Zero-DCE [31]	0.6322	15.2961	0.2925
SCI [43]	0.6514	16.4265	0.3156
proposed	**0.7963**	**20.7127**	0.2271

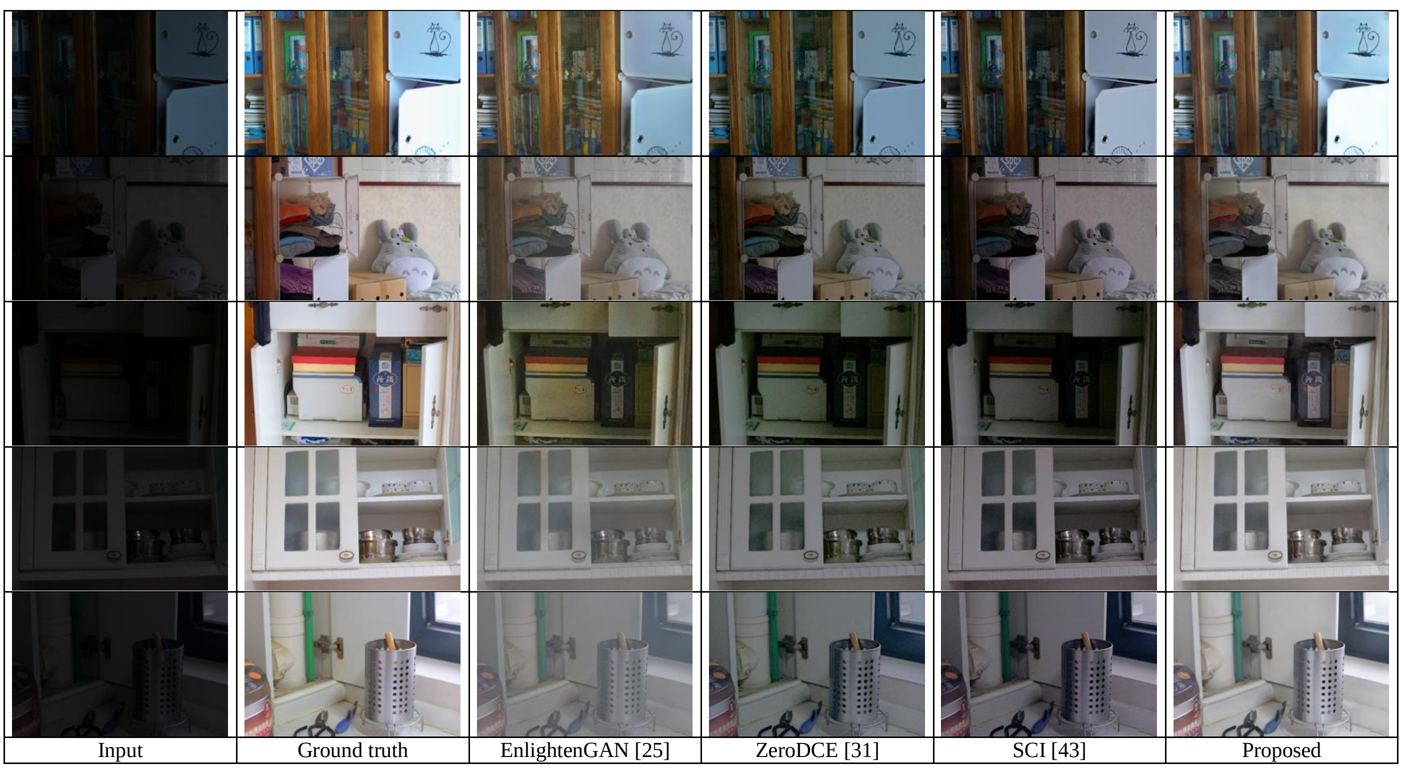

Fig. 5. The enhanced results from the LOL-dataset by different methods.

These results demonstrate the effectiveness of the proposed illumination attention mechanism and edge-aware optimization in guiding the network to enhance low-light images with improved structure, texture, and perceptual realism.

B. Ablation Study

To investigate the efficiency of each module in the proposed model to the final results, four experiments were designed. The following ablated Components were combined for training: U-Net as generator, PatchGAN as discriminator, and Attention module. Finally, the generated models were compared to determine the role of each module in the model. From the Table II, it can be observed:

1) Removing the illumination attention results in PSNR and SSIM decrease, and a deterioration in the LPIPS metric, indicating that illumination perception significantly enhances detail and realism.

2) After removing the PatchGAN discriminator, the image quality metrics slightly decline, suggesting that adversarial training also contributes to improving local details.

3) Without using U-net as the generator, the overall performance decreases, demonstrating that the U-net encoder-decoder structure plays a crucial role in feature modelling under complex illumination conditions.

TABLE II EVALUATION RESULTS FOR ABLATION STUDIES ON LOL-DATASET

Components	SSIM↑	PSNR↑	LPIPS ↓
Full model	**0.7963**	**20.7127**	**0.2271**
w/o attention	0.7012	16.4131	0.3422
w/o patchGAN	0.7054	18.0921	0.3012
w/o U-net	0.5023	12.1091	0.3547

VI. CONCLUSION

This paper addresses the existing challenges in the LLIE task, such as color distortion, noise amplification, and detail loss, and proposes a novel BA-GAN. This framework aims to generate enhanced images that are visually appealing and have high information fidelity through a well-designed end-to-end deep learning model.

The core contribution of this paper lies in the hybrid architecture design of BA-GAN. This paper adopts a U-Net-based generator, which effectively captures multi-scale context features and preserves fine image details through skip connections. The key innovation is that this paper integrates a novel illumination attention module into the decoder of the generator, enabling it to adaptively focus on the regions that require key illumination correction according to the image content. Meanwhile, this paper uses the PatchGAN discriminator architecture to ensure the local realism of the generated images. The entire model is optimized through a composite loss function that combines adversarial loss, brightness attention loss, and perceptual loss, thus achieving an effective balance among the realism, brightness accuracy, and structural fidelity of the images.

To verify the effectiveness of our method, this paper conducted a series of extensive experiments on the public LOL dataset. The quantitative evaluation results show that our BA-GAN outperforms a variety of current mainstream methods in terms of PSNR (20.7127) and SSIM (0.7963), demonstrating its superior ability in structure preservation and noise suppression. Qualitative visual comparisons further prove that our method can significantly improve the visibility of images while effectively suppressing noise and maintaining natural color features. Ablation experiments also verify the key contributions of each module this paper proposed to the final performance.

Although this study has achieved remarkable results, there are still some limitations and future exploration directions. First, as a supervised learning method, our model currently relies on paired training data, which is difficult to obtain in many real-world scenarios. Future work can explore combining the illumination attention mechanism with unsupervised or semi-supervised learning frameworks to enhance the model's generalization ability. Second, further research can be conducted to extend this method from static images to the field of low-light video enhancement and explore how to ensure the temporal consistency between video frames. Finally, future research can systematically evaluate the actual improvement effect of the images enhanced by BA-GAN on the performance of downstream computer vision tasks (such as object detection and image classification).

REFERENCES

[1] Y. P. Loh and C. S. Chan, “Unveiling contrast in darkness,” in 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), IEEE, 2015, pp. 266–270. doi: 10.1109/ACPR.2015.7486507.

[2] W. Wang, X. Wu, X. Yuan, and Z. Gao, “An experiment-based review of low-light image enhancement methods,” Ieee Access, vol. 8, pp. 87884–87917, 2020.

[3] M. Leo, G. Medioni, M. Trivedi, T. Kanade, and G. M. Farinella, “Computer vision for assistive technologies,” Computer Vision and Image Understanding, vol. 154, pp. 1–15, 2017, doi: 10.1016/j.cviu.2016.09.001.

[4] C. Li et al., “Low-light image and video enhancement using deep learning: A survey,” IEEE transactions on pattern analysis and machine intelligence, vol. 44, no. 12, pp. 9396–9416, 2021, doi: 10.1109/TPAMI.2021.3126387.

[5] R. Al Sobbahi and J. Tekli, “Low-light homomorphic filtering network for integrating image enhancement and classification,” Signal Processing: Image Communication, vol. 100, p. 116527, 2022, doi: 10.1016/j.image.2021.116527.

[6] R. C. Gonzalez, Digital image processing. Pearson education india, 2009.

[7] E. H. Land and J. J. McCann, “Lightness and retinex theory,” Journal of the Optical society of America, vol. 61, no. 1, pp. 1–11, 1971, doi: 10.1364/JOSA.61.000001.

[8] Z. He et al., “Low-light image enhancement with multi-scale attention and frequency-domain optimization,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 34, no. 4, pp. 2861–2875, 2023.

[9] I. J. Goodfellow et al., “Generative adversarial networks,” Communications of the ACM, vol. 63, pp. 139–144, 2014.

[10] A. Madbouly, “Increase contrast of low light image using modified histogram equalization,” Advances in Basic and Applied Sciences, vol. 3, no. 1, pp. 67–71, 2024, doi: 10.21608/abas.2024.327925.1052.

[11] B. S. Rao, “Dynamic histogram equalization for contrast enhancement for digital images,” Applied Soft Computing, vol. 89, p. 106114, 2020, doi: 10.1016/j.asoc.2020.106114.

[12] D. J. Jobson, Z. Rahman, and G. A. Woodell, “Properties and performance of a center/surround retinex,” IEEE transactions on image processing, vol. 6, no. 3, pp. 451–462, 1997.

[13] D. J. Jobson, Z. 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, vol. 6, no. 7, pp. 965–976, 1997, doi: 10.1109/83.597272.

[14] H. Liu, X. Sun, H. Han, and W. Cao, “Low-light video image enhancement based on multiscale Retinex-like algorithm,” 2016 Chinese Control and Decision Conference (CCDC), pp. 3712–3715, 2016, doi: 10.1109/CCDC.2016.7531629.

[15] S. Parthasarathy and P. Sankaran, “An automated multi scale retinex with color restoration for image enhancement,” presented at the 2012 National Conference on Communications (NCC), IEEE, 2012, pp. 1–5.

[16] 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, doi: 10.1016/j.patcog.2016.06.008.

[17] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, “Msr-net: Low-light image enhancement using deep convolutional network,” arXiv preprint arXiv:1711.02488, 2017.

[18] 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 Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 5901–5910.

[19] Y. Zhang, 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, pp. 1632–1640.

[20] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang, “Beyond Brightening Low-light Images,” International Journal of Computer Vision, vol. 129, pp. 1013–1037, 2021.

[21] L. Chang, G. Zhou, O. Soufan, and J. Xia, “miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology,” Nucleic Acids Research, vol. 48, pp. W244–W251, 2020, doi: 10.1093/nar/gkaa467.

[22] Y. Liu, T. Huang, W. Dong, F. Wu, X. Li, and G. Shi, “Low-Light Image Enhancement with Multi-stage Residue Quantization and Brightness-aware Attention,” 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 12106–12115, 2023, doi: 10.1109/ICCV51070.2023.01115.

[23] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” ArXiv, vol. abs/1505.04597, 2015, [Online]. Available: https://api.semanticscholar.org/CorpusID:3719281

[24] C. C. Lim, Y. P. Loh, and L. Wong, “LAU-Net: A low light image enhancer with attention and resizing mechanisms,” Signal Process. Image Commun., vol. 115, p. 116971, 2023, doi: 10.1016/j.image.2023.116971.

[25] Y. Jiang et al., “EnlightenGAN: Deep Light Enhancement Without Paired Supervision,” IEEE Transactions on Image Processing, vol. 30, pp. 2340–2349, 2019, doi: 10.1109/TIP.2021.3051462.

[26] T. Ma et al., “RetinexGAN:Unsupervised Low-Light Enhancement With Two-Layer Convolutional Decomposition Networks,” IEEE Access, vol. 9, pp. 56539–56550, 2021, doi: 10.1109/ACCESS.2021.3072331.

[27] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976, 2016.

[28] J. Hu, X. Guo, J. Chen, G. Liang, F. Deng, and T. L. Lam, “A Two-Stage Unsupervised Approach for Low Light Image Enhancement,” IEEE Robotics and Automation Letters, vol. 6, pp. 8363–8370, 2020, doi: 10.1109/LRA.2020.3048667.

[29] R.-C. Wang, B. Jiang, C. Yang, Q. Li, and B. Zhang, “MAGAN: Unsupervised low-light image enhancement guided by mixed-attention,” Big Data Min. Anal., vol. 5, pp. 110–119, 2022, doi: 10.26599/BDMA.2021.9020020.

[30] W. Kozlowski, M. Szachniewicz, M. Stypulkowski, and M. Zieba, “Dimma: Semi-Supervised Low-Light Image Enhancement with Adaptive Dimming,” Entropy, vol. 26, 2023, [Online]. Available: https://api.semanticscholar.org/CorpusID:264146400

[31] C. Guo et al., “Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement,” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1777–1786, 2020, doi: 10.1109/CVPR42600.2020.00185.

[32] A. Zhu, L. Zhang, Y. Shen, Y. Ma, S. Zhao, and Y. Zhou, “Zero-Shot Restoration of Underexposed Images via Robust Retinex Decomposition,” 2020 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6, 2020.

[33] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, “Retinex-inspired Unrolling with Cooperative Prior Architecture Search for Low-light Image Enhancement,” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10556–10565, 2020.

[34] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.

[35] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” ArXiv, vol. abs/1603.08155, 2016, [Online]. Available: https://api.semanticscholar.org/CorpusID:980236

[36] G. Zhang, Q. Li, Q. Li, and T. Wu, “Low-light image enhancement model based on self-calibration illumination and color adjustment,” 2022 2nd International Conference on Algorithms, High Performance Computing and Artificial Intelligence (AHPCAI), pp. 559–562, 2022.

[37] D. D. Jonge, “LuminanceL1Loss: A loss function which measures percieved brightness and colour differences,” ArXiv, vol. abs/2311.04614, 2023, [Online]. Available: https://api.semanticscholar.org/CorpusID:265050838

[38] Y. Atoum, M. Ye, L. Ren, Y. Tai, and X. Liu, “Color-wise Attention Network for Low-light Image Enhancement,” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 2130–2139, 2019.

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

[40] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Transactions on Image Processing, vol. 13, pp. 600–612, 2004, doi: 10.1109/TIP.2003.819861.

[41] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 586–595, 2018, doi: 10.1109/CVPR.2018.00068.

[42] J. Guo, J. Ma, Á. F. García-Fernández, Y. Zhang, and H. Liang, “A survey on image enhancement for Low-light images,” Heliyon, vol. 9, no. 4, 2023.

[43] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, “Toward fast, flexible, and robust low-light image enhancement,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 5637–5646.