# Low-Light Image Enhancement Using Gamma Learning And Attention-Enabled Encoder-Decoder Networks

Bibhabasu Debnath, Sahana Ray, and Sanjay Ghosh, *Senior Member, IEEE,*

***Abstract*—Images acquired in low-light environments present significant obstacles for computer vision systems and human perception, especially for applications requiring accurate object recognition and scene analysis. Such images typically manifest multiple quality issues: amplified noise, inadequate scene illumination, contrast reduction, color distortion, and loss of details. While recent deep learning methods have shown promise, developing simple and efficient frameworks that naturally integrate global illumination adjustment with local detail refinement continues to be an important objective. To this end, we introduce a dual-stage deep learning architecture that combines adaptive gamma correction with attention-enhanced refinement to address these fundamental limitations. The first stage uses an Adaptive Gamma Correction Module (AGCM) to learn suitable gamma values for each pixel based on both local and global cues, producing a brightened intermediate output. The second stage applies an encoder-decoder deep network with Convolutional Block Attention Modules (CBAM) to this brightened image, in order to restore finer details. We train the network using a composite loss that includes L1 reconstruction, SSIM, total variation, color constancy, and gamma regularization terms to balance pixel accuracy with visual quality. Experiments on LOL-v1, LOL-v2 real, and LOL-v2 synthetic datasets show our method reaches PSNR of upto 29.96 dB and upto 0.9458 SSIM, outperforming existing approaches. Additional tests on DICM, LIME, MEF, and NPE datasets using NIQE, BRISQUE, and UNIQUE metrics confirm better perceptual quality with fewer artifacts, achieving the best NIQE scores across all datasets. Our `GAtED` (Gamma learned and Attention-enabled Encoder-Decoder) method effectively handles both global illumination adjustment and local detail enhancement, offering a practical solution for low-light enhancement. The code should be accessible at https://github.com/bibhabasuiitkgp/GAtED.**

***Index Terms*—Low-light image enhancement, learnable gamma correction, attention mechanism, U-Net, deep learning.**

## I. Introduction

Images captured in poor lighting conditions pose substantial challenges for both human observation and computer vision applications, particularly in object detection tasks [1]–[3]. When optical devices operate under inadequate illumination or when external lighting varies significantly, the resulting photographs exhibit multiple degradation artifacts: amplified noise, insufficient illumination, reduced contrast, color distortion, and loss of fine details. Manual correction of these images proves both time-intensive and frequently yields suboptimal results. Consequently, researchers have developed numerous low-light image enhancement (LLIE) algorithms to tackle these degradation problems and improve image quality.

This work is supported by the Faculty Start-up Research Grant (FSRG), Indian Institute of Technology Kharagpur, India.

Bibhabasu Debnath, Sahana Ray, and Sanjay Ghosh are with the Department of Electrical Engineering, Indian Institute of Technology Kharagpur, WB 721302, India (email: sanjay.ghosh@ee.iitkgp.ac.in).

Existing LLIE approaches fall into two broad categories: traditional methods and deep learning frameworks. Traditional methods build upon classical image processing principles such as histogram equalization [4], [5], Retinex-based methods [6] and optimization strategies [7], [8]. While [5] improves contrast by adaptively redistributing pixel intensities, [8] refine illumination maps through structure-aware enhancement. These conventional approaches offer computational efficiency and mathematical interpretability; however, they often suffer from color distortion and edge artifacts, and typically rely on restrictive assumptions about illumination distributions.

Deep learning-based approaches have achieved superior performance through data-driven learning. Within this category, Retinex-inspired models [9]–[18] decompose images into illumination and reflectance components using trainable networks. KinD [11] employs dedicated subnetworks for reflectance restoration and illumination enhancement, while Retinex-Net [9] introduces trainable Decom-Net followed by encoder-decoder enhancement. Advanced frameworks like URetinex-Net++ [12] integrate implicit prior regularization and cross-stage fusion blocks for improved noise suppression and detail preservation. Retinexformer [15] leverages Illumination-Guided Transformers to model long-range dependencies and handle varying lighting conditions adaptively. CUE [16] presents a Retinex-based Customized Unfolding Enhancer with Masked Autoencoder-based priors, while Diff-Retinex++ [17] integrates diffusion models with Retinex-driven restoration for physically-inspired generative enhancement. Among supervised CNN-based algorithms, [19] regards low-light enhancement as a residual learning problem, while [20] uses nested skip connections and residual blocks based on UNet++. On the other hand, unsupervised frameworks such as Zero-DCE [21] predicts high-order curves for pixel-wise adjustment without paired data, while EnlightenGAN [22] leverages a lightweight GAN architecture to learn unpaired mappings from low- to normal-light images. More recently, diffusion-based frameworks [17], [23]–[27] have emerged as powerful approaches for image restoration and enhancement. PyDiff [24] employs progressive pyramid-resolution sampling strategies with global correctors for generalization to unseen noise distributions. DiffUIR [25] proposes selective hourglass mapping to handle diverse degradation distributions simultaneously. GSAD [26] introduces uncertainty-guided regularization for enhanced detail and noise suppression. Despite their generative capabilities, these models require significant computational resources and multiple sampling steps, limiting real-time applicability. Contemporary LLIE research has increasingly incorporated attention mechanisms [28] and adaptive gamma correction to handle non-uniform illumination and multi-scale

degradations [29]–[31]. Yang et al. [32] proposed a Retinex-based framework incorporating pixel-wise gamma maps following image decomposition, learning adaptive corrections for uneven lighting while preserving structure. Wang et al. [33] introduced hierarchical gamma maps learned through Taylor expansion coupled with transformer-based global modeling, accelerating inference while maintaining adaptive illumination adjustment.

In this work, we propose a dual-stage enhancement architecture that leverages an Adaptive Gamma Correction Module (AGCM) followed by a U-Net augmented by a convolutional block attention module (CBAM). The first stage (AGCM) computes a gamma correction parameter map and applies a power law correction. The second stage (U-Net augmented CBAM) is a U-Net architecture integrated with Convolutional Block Attention Modules (CBAM) that learns to enhance contrast, restore the edge, and texture restoration, and noise suppression. The motivation behind the dual-stage enhancement architecture is that global illumination simplifies the task for the U-net to recover textures and colors.

The key contributions in our dual-stage method are:

- Adaptive Gamma Correction Metwork (AGCM) for illumination recovery of the underexposed image by applying a differentiable learned gamma correction map. The gamma Correction map estimates pixel-wise gamma values to adaptively brighten different regions based on content. But while boosting brightness and improving visibility, it fails to restore the fine textures, Noise suppression, and color balancing. To address the above problem, the CBAM-enhanced U-Net architecture is introduced after the gamma-corrected sample.
- Attention-enhanced U-Net for local detail recovery and perceptual quality. This is designed to refine the gamma-corrected images by restoring lost textures, suppressing noise, and color imbalance. The architecture integrates Convolutional Block Attention Modules (CBAM) within a U-Net framework to dynamically focus on salient spatial regions and informative channels during both encoding and decoding. While the gamma-corrected image offers a globally brightened view, it often lacks structural sharpness and perceptual consistency. The attention-enhanced U-Net addresses these limitations by leveraging skip connections, multi-scale features, and attention-guided refinement to recover fine details and enhance overall image realism.
- Extensive experiments are performed on LOL-v1 and LOL-v2 (Real-Captured) datasets, comparing the proposed method with several of its state-of-the-art alternatives. Results demonstrate the superiority of our proposed **GAtED** (**G**amma Learned and **At**tention-enabled **E**ncoder-**D**ecoder) method.

The rest of this paper follows a systematic presentation of our contributions and their significance. In Section II, the existing literature on LLIE is explored, covering both traditional methods and contemporary deep learning approaches, with particular emphasis on attention mechanisms and adaptive gamma correction techniques. Section III details our proposed GAED methodology, including its architecture, mathematical formulations, and the expression of the multi-term loss function. Section IV outlines our experimental framework, introducing the datasets used, the metrics evaluated, and the implementation specifications. Additionally, it highlights the superior performance achieved by our proposed method compared to its state-of-the art alternatives, through qualitative and quantitative evaluations. Section V concludes with a discussion of the significance and achievements of the contributions proposed in this work.

## II. Related Work

Over the past decades, low-light image enhancement (LLIE) has received significant research attention, resulting in a variety of traditional and deep learning-based approaches.

### A. Traditional LLIE Methods

*1) Histogram-Based Methods:* Histogram Equalization (HE) [4] improves image contrast by redistributing pixel intensity values but often results in over-enhancement and noise amplification, particularly in extremely dark or bright regions. Dynamic Histogram Equalization (DHE) [5] addresses this issue by dividing the histogram into sub-histograms based on local minima and allocating adaptive dynamic gray-level ranges, thus preserving local details and preventing dominant regions from overwhelming weaker features.

*2) Retinex-Based Methods:* Retinex theory [34], [35] models an image as a product of illumination and reflectance. Since illumination is not directly observable, it must be estimated through various strategies, such as passing the image through filters [6], [36]. Single-Scale Retinex (SSR) and Multi-Scale Retinex (MSR) [37], [38] enhance visibility by improving the illumination component using logarithmic transformations.

*3) Optimization-Based Methods:* LIME [8] estimates an initial illumination map by taking the pixel-wise maximum across RGB channels and refines it via an optimization method, producing structure-aware enhancement. LIME's use of the max-RGB heuristic may lead to inaccurate illumination estimation under colored lighting and introduces edge noise, limiting its robustness.

### B. Deep Learning-Based LLIE Methods

*1) Retinex-Inspired Models:* KinD [11] designs two dedicated subnetworks for reflectance restoration and illumination enhancement, helping preserve structural integrity and reduce color distortion. Retinex-Net [9] introduces a trainable Decom-Net to decompose the image into illumination and reflectance, followed by an Enhance-Net to boost illumination using an encoder-decoder structure. URetinex-Net [10] unfolds an optimization-based Retinex formulation into a trainable network and integrates a learnable illumination adjustment module to better adapt to challenging lighting conditions. URetinex-Net++ [12] further enhances this framework by introducing implicit prior regularization and a cross-stage fusion block, enabling adaptive decomposition with improved noise suppression, detail preservation, and color fidelity. RUAS [13]

builds a lightweight Retinex-inspired unrolled network and employs a cooperative reference-free learning strategy to discover effective low-light prior architectures, enabling fast and resource-efficient image enhancement. CRetinex [14] introduces a color-shift aware Retinex model that decomposes images into reflectance, color shift, and illumination components, enabling improved color constancy. Retinexformer [15] addresses hidden corruptions in low-light images through a one-stage Retinex-based framework that employs an Illumination-Guided Transformer to model long-range dependencies and adaptively handle varying lighting conditions. CUE [16] presents a Retinex-based Customized Unfolding Enhancer that leverages Masked Autoencoder-based illumination and noise priors within structure and optimization flows, enabling more effective illumination restoration and noise suppression in low-light image enhancement. Diff-Retinex++ [17] introduces a Retinex-driven reinforced diffusion model that integrates a diffusion model with Retinex-driven restoration, with the aim of achieving physically-inspired generative enhancement. Retinex-Inspired Reconstruction Optimization (RIRO) model treats low-light enhancement as an image reconstruction task using spatial and frequency domain features [18].

*2) CNN-Based Models:* In [19], low-light enhancement is treated as a residual learning problem and uses a Deep Lightening Network (DLN) to estimate the residual between normal-light and low-light images. On the other hand, [20] enhances low-light images using a UNet++ architecture with nested skip connections and Instance Normalization-based residual blocks. Zero-DCE [21] takes a different approach by predicting high-order curves for pixel-wise adjustment on the input image in an unsupervised manner using a Deep Curve Estimation Network. EnlightenGAN [22] leverages a lightweight, one-path GAN architecture to learn an unpaired mapping from low- to normal-light images without relying on paired data, enabling faster and more flexible training.

*3) Diffusion-Based Frameworks:* Diffusion models like [17] have recently emerged as a powerful approach to image restoration and enhancement. The Controllable Light Enhancement (CLE) Diffusion Model [23] introduces an illumination embedding for user-adjustable brightness and integrates the Segment-Anything Model (SAM) for user-defined region-specific enhancement. Pyramid Diffusion Model (PyDiff) employs a progressive pyramid-resolution sampling strategy and incorporates a global corrector to make the model faster and generalizable to unseen noise and illumination distributions [24]. DiffUIR framework [25] proposes a selective hourglass mapping strategy and shared distribution term (SDT), enabling the model to handle different degradation distributions at once. GSAD [26] introduces uncertainty-guided regularization to enhance detail, contrast, and noise suppression in low-light images. In [27], a zero-reference low-light enhancement framework is proposed, leveraging physical quadruple priors and a pretrained generative diffusion model.

*4) Transformers and Attention Mechanisms:* The attention mechanism was initially proposed for visual processing [39]. It allows neural networks to focus on important features while suppressing irrelevant information [40]. Dual attention modules like Convolutional Block Attention Module (CBAM) [28] combine both types sequentially. In LLIE, attention helps to identify underexposed regions, suppress noise, and improve structure. [21] and [41] used attention to enhance brightness and detail adaptively. Restormer [29] redesigns the multi-head attention and feed-forward network blocks to capture long-range pixel interactions. LLFormer [30] featured axis-based multi-head self-attention and cross-layer attention fusion. NeRCo [31] employs an implicit neural representation with semantic-oriented supervision to robustly recover perceptual-friendly visual results.

*5) Adaptive Gamma Correction-Based Methods::* Yang et al. [32] proposed a lightweight Retinex-based deep framework that incorporates an adaptive gamma correction module following image decomposition into illumination and reflectance components. Wang et al. [33] further introduced an illumination-aware gamma correction network, where hierarchical gamma maps are learned and approximated via Taylor expansion to accelerate inference, coupled with a transformer-based global modeling module. Unlike traditional approaches, these methods adaptively infer gamma values instead of using hand-crafted histograms or fixed curves.

## III. Proposed Method

In this section, we present the complete details of our proposed deep learning method. We first introduce the proposed two-stage network architecture. This is followed by the details of our proposed loss function to guide the training of the proposed deep learning method. The first stage of the network, shown in Figure 1, consists of an enhancement module that basically produces an enhanced image by performing a spatially adaptive gamma correction operation. In the second stage, we further enhance the low-light image by integrating a convolutional block attention module (CBAM) into an UNet framework, applying both channel and spatial attention sequentially, as shown in Figure 2.

### A. Network Architecture

*1) Adaptive Gamma Correction Module (AGCM):* This module aims to produce an initial enhancement of the low-light image by adjusting illumination in a spatially adaptive manner. Traditional gamma correction applies a fixed exponent to all pixels, which fails to address non-uniform illlumination. To address this, our AGCM predicts a pixel-wise gamma map that adapts the gamma value based on both local texture and global context. The AGCM consists of a hierarchical *Feature Extraction Block (FEB)* followed by a *Global Context Block (GCB)*.

*a) Feature Extraction Block (FEB):* The FEB captures local structures and textures of the input image. It uses three successive $3 \times 3$ convolution layers, each followed by batch normalization and ReLU activation, as shown in Figure 1. Each convolution extracts feature responses at increasing levels of abstraction, i.e., edges, corners, and local intensity patterns. For an input low-light image $I_{\text{in}} \in \mathbb{R}^{3\times H\times W}$, we obtain the deep feature representation $F_3 \in \mathbb{R}^{32\times H\times W}$, which encodes local information that determines how much each pixel should be enhanced.

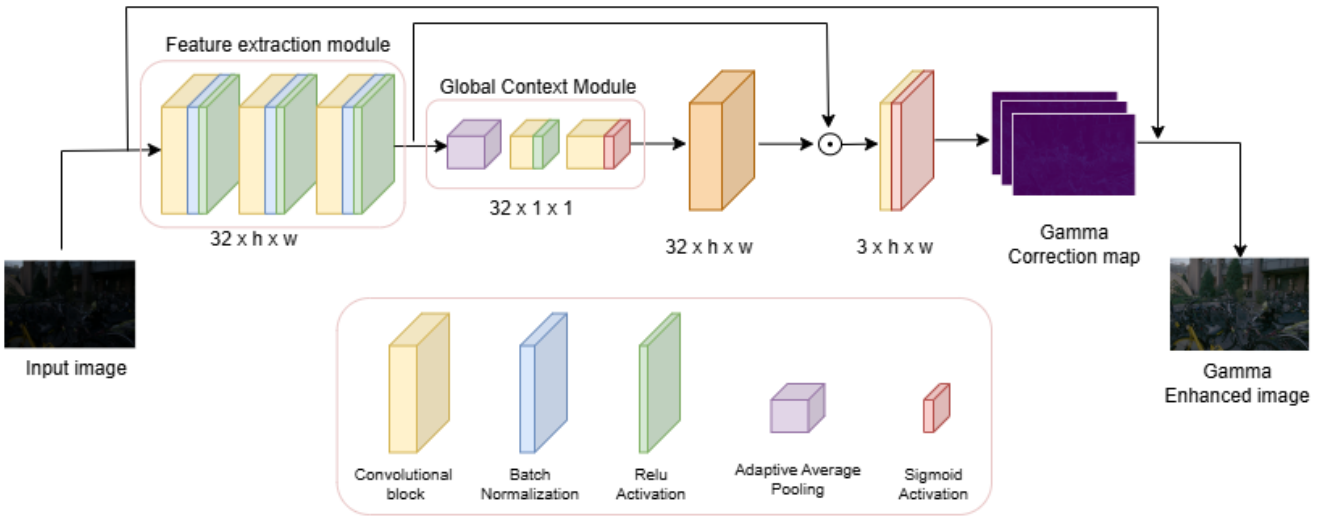


Fig. 1: Adaptive Gamma Correction Module (AGCM): the first stage of our proposed dual-stage deep learning method for low-light enhancement. It first extracts local textures and then injects global luminance cues to learn pixel-wise gamma for content-aware brightening. Note that we restricted the $\gamma$ values in $[0.5, 2.0]$ to avoid over-exposed while also preserving well-lit regions in the image.

*b) Global Context Block (GCB):* While local features describe textures and edges, they lack awareness of the overall illumination and contrast distribution across the entire image. The GCB addresses this by integrating global context through an adaptive pooling and attention mechanism. First, adaptive average pooling condenses spatial information into a global descriptor $G \in \mathbb{R}^{32\times1\times1}$, representing the overall luminance statistics. Then, two successive $1\times1$ convolutions, interleaved with ReLU and Sigmoid activations, model non-linear relationships between channels:

$$C_1 = \text{ReLU}(\text{Conv}^{16}_{1\times1}(G)), \quad C_2 = \sigma(\text{Conv}^{32}_{1\times1}(C_1)). \quad (1)$$

Here, $C_2$ acts as a channel attention map indicating the importance of each feature channel in the local feature map $F_3$. Multiplying $C_2$ with $F_3$ modulates the response of each channel according to global illumination cues:

$$F_{\text{ctx}} = F_3 \odot C_2, \quad (2)$$

where $\odot$ represents point-wise multiplication. This step allows darker or underexposed regions to receive stronger activations, enabling more effective enhancement later.

*c) Pixel-wise Gamma Prediction and Correction:* After contextual refinement, the module predicts a **gamma correction map** that determines how much to brighten or darken each pixel. A $1\times1$ convolution followed by a scaled Sigmoid activation produces this gamma map:

$$\gamma = 0.5 + 1.5\,\sigma(\text{Conv}^{3}_{1\times1}(F_{\text{ctx}})), \quad (3)$$

where $\sigma$ denotes the sigmoid activation function applied after a $1\times1$ convolution with 3 output channels on the context-enhanced features $F_{\text{ctx}}$, scaled and shifted to map outputs to the range $[0.5, 2.0]$. The $\sigma$ nonlinearity operation ensures all predicted gamma values lie between 0 and 1, and the scaling extends this range to $[0.5, 2.0]$. Pixels with $\gamma < 1$ are brightened (since raising to a power less than 1 increases intensity), while those with $\gamma > 1$ are slightly darkened to prevent overexposure. Finally, the enhanced image is computed by applying pixel-wise gamma correction:

$$I_\gamma(x,y) = (I_l(x,y) + \varepsilon)^{\gamma(x,y)}, \quad \varepsilon = 10^{-6}, \quad (4)$$

where $I_l(x,y)$ represents the low-light input image intensity at pixel location $(x,y)$, raised to the power of the predicted gamma value $\gamma(x,y)$ at that location, with a small constant $\varepsilon$ added to prevent numerical instability for near-zero pixel values. This equation performs a content-aware illumination mapping - darker pixels are brightened more aggressively, while brighter regions are preserved. Through this adaptive mechanism, the AGCM produces a balanced, illumination-corrected version of the input image, serving as the initial enhanced output for the next stage.

*2) Attention-enabled Encoder–Decoder Architecture:* Although AGCM improves illumination adaptively, it may not fully restore lost details or textures, especially in very dark regions. Hence, the second stage leverages an attention-augmented UNet architecture to refine and reconstruct the image with better contrast, structural consistency, and color balance.

*a) Convolutional Block Attention Module (CBAM):* To strengthen the model's ability to focus on relevant features, we integrate the Convolutional Block Attention Module (CBAM) into the UNet encoder and decoder blocks. CBAM sequentially applies channel attention and spatial attention, helping the network to understand what and where to improve.

*Channel Attention:* Each feature map $F \in \mathbb{R}^{C\times H\times W}$ is summarized using global average pooling and global max pooling, producing two channel descriptors ($F_{\text{avg}}$ and $F_{\text{max}}$). These pass through a shared two-layer MLP to capture non-

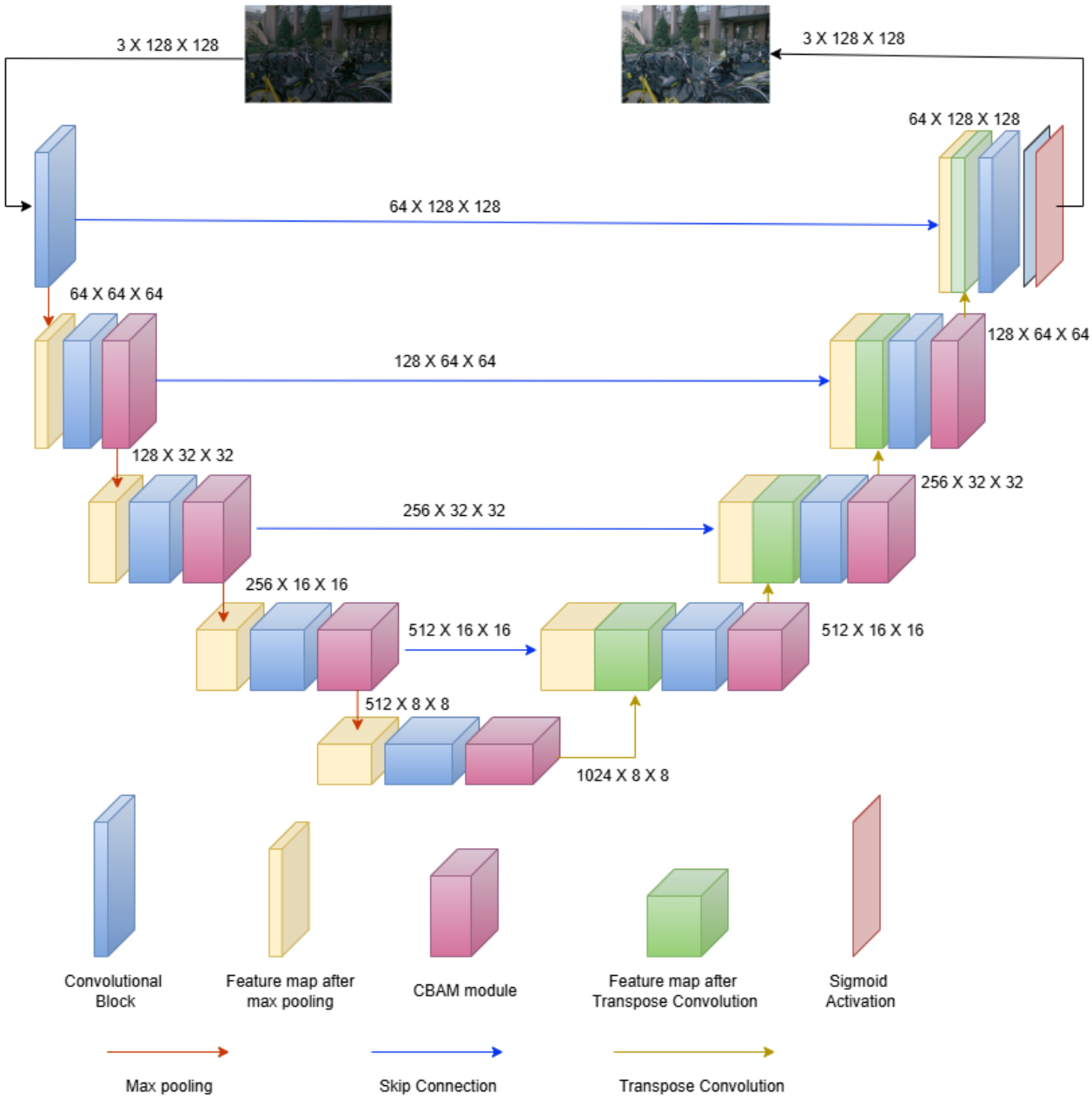


Fig. 2: CBAM-enhanced U-Net for refinement. Channel attention amplifies informative features, and spatial attention localises salient regions for enhancement. Skip connections with upsampling restore details while suppressing noise and colour imbalance. A final 1×1 convolution maps refined features to the enhanced image.

linear dependencies between channels. The resulting channel attention map $M_c$ highlights the most informative feature channels:

$$M_{\mathrm{c}} = \sigma(\mathrm{MLP}(F_{\mathrm{avg}}) + \mathrm{MLP}(F_{\mathrm{max}})),$$

where $F_{\mathrm{avg}}$ and $F_{\mathrm{max}}$ are the globally averaged and max-pooled channel descriptors, each processed through a shared multi-layer perceptron (MLP), summed element-wise, and passed through a sigmoid activation $\sigma$ to produce channel-wise attention weights. Multiplying $M_c$ with $F$ amplifies globally useful features (e.g., edges and texture cues) and suppresses irrelevant noise.

*Spatial Attention:* After channel attention, spatial attention determines which regions of the image are most important. It compresses the refined feature map $F'$ along the channel dimension using average and max pooling:

$$\begin{aligned} F'_{\mathrm{avg}} &= \mathrm{Mean}_{\mathrm{chan}}(F'), \\ F'_{\mathrm{max}} &= \mathrm{Max}_{\mathrm{chan}}(F'), \\ F_{\mathrm{cat}} &= \mathrm{Concat}(F'_{\mathrm{avg}}, F'_{\mathrm{max}}), \\ M_{\mathrm{s}} &= \sigma(\mathrm{Conv}_{7\times7}(F_{\mathrm{cat}})), \end{aligned}$$

where $F'_{\mathrm{avg}}$ and $F'_{\mathrm{max}}$ represent the channel-wise mean and maximum operations applied to the channel-refined feature

map $F'$, which are then concatenated along the channel dimension to form $F_{\text{cat}}$, followed by a $7 \times 7$ convolution and sigmoid activation to generate the spatial attention map $M_{\text{s}}$. This map assigns higher weights to spatial regions that are more relevant for enhancement. The final refined output is obtained as:

$$F_{\text{enhanced}} = M_{\text{s}} \odot F', \tag{5}$$

where $\odot$ denotes element-wise multiplication between the spatial attention map $M_{\text{s}}$ and the channel-refined features $F'$. Together, CBAM ensures that each feature map is selectively refined in both the channel and spatial dimensions, helping the UNet focus its enhancement efforts where they matter most.

*b) Encoder–Decoder Framework:* The attention-augmented UNet consists of symmetric encoder and decoder paths, each built from Double Convolution Blocks. Each block contains two consecutive $3 \times 3$ convolutions with batch normalization and ReLU activations:

$$F_1 = \text{ReLU}(\text{BN}(\text{Conv}_{3\times3}(x))),$$
$$F_2 = \text{ReLU}(\text{BN}(\text{Conv}_{3\times3}(F_1))),$$

where the input $x$ undergoes a $3 \times 3$ convolution, batch normalization (BN), and ReLU activation to produce $F_1$, which is then processed through another identical convolution-normalization-activation sequence to yield $F_2$. The encoder progressively reduces spatial dimensions via max-pooling, capturing hierarchical features. Each downsampling step is wrapped in a CBAM to refine the feature map before deeper encoding:

$$D_i = \text{DoubleConv}(\text{CBAM}(\text{MaxPool}(F_{i-1}))),$$

where $i = 1, 2, 3, 4$ and the feature map $F_{i-1}$ from the previous layer is first max-pooled to reduce spatial resolution, then refined through CBAM, and finally processed through a double convolution block to produce the encoder output $D_i$. At the bottleneck, the representation $F_B \in \mathbb{R}^{1024\times8\times8}$ encodes both global structure and fine details.

The decoder then gradually restores spatial resolution through transposed convolutions. Skip connections from corresponding encoder layers preserve fine details, while CBAM modules guide attention during upsampling:

$$U_i = \text{DoubleConv}(\text{CBAM}(\text{Concat}(\text{UpConv}(U_{i+1}), D_i))),$$

where $i = 3, 2, 1, 0$ and the decoder output $U_{i+1}$ from the previous layer is upsampled via transposed convolution (UpConv), concatenated with the corresponding encoder feature map $D_i$ from the skip connection, refined through CBAM, and processed by a double convolution block to produce $U_i$. Finally, a $1 \times 1$ convolution followed by a Sigmoid activation maps the decoder output to the final enhanced image:

$$\hat{x} = \sigma(\text{Conv}_{1\times1}(U_0)) \in \mathbb{R}^{3\times128\times128}, \tag{6}$$

where the final decoder output $U_0$ is passed through a $1 \times 1$ convolution to reduce channels to 3 (RGB), followed by sigmoid activation $\sigma$ to constrain pixel values to $[0, 1]$, producing the enhanced image $\hat{x}$. This output combines the global brightness correction from the first stage with fine-grained restoration and detail enhancement from the second stage, yielding a visually balanced, high-quality result.

### B. Loss Function

*1) L1 Reconstruction Loss:* We employ the *L1 Reconstruction Loss* to ensure that the enhanced image remains faithful to the ground truth at the pixel level. This loss penalizes the absolute differences between the predicted and target pixel intensities as follows.

$$\mathcal{L}_{\text{L1}} = \frac{1}{CHW} \sum_{c=1}^{C} \sum_{i=1}^{H} \sum_{j=1}^{W} |\hat{x}_{c,i,j} - x_{c,i,j}| \tag{7}$$

*2) SSIM Loss:* In order to preserve structural similarity and perceptual quality in the output image, we define the structural similarity index measure (SSIM) loss. The SSIM metric between the enhanced image and the ground truth image is computed as:

$$\text{SSIM}(x, \hat{x}) = \frac{(2\mu_1\mu_2 + C_1)(2\sigma_{12} + C_2)}{(\mu_1^2 + \mu_2^2 + C_1)(\sigma_1^2 + \sigma_2^2 + C_2)} \tag{8}$$

where $\mu_1$ and $\mu_2$ are the pixel sample means of the images; $\sigma_1$ and $\sigma_2$ are the sample variances of the images; $\sigma_{12}$ the sample covariance. Here, $C_1$ and $C_2$ are constants set to $0.01^2$ and $0.03^2$, respectively, to stabilize the division. Finally, the SSIM loss is

$$\mathcal{L}_{\text{SSIM}} = 1 - \text{SSIM}(x, \hat{x}).$$

*3) Total Variation (TV) Loss:* We have a loss component based on the total variation (TV) of the enhanced/output image to penalize abrupt intensity changes between neighboring pixels, promoting spatial smoothness while preserving important edges. We compute TV loss as follows: We compute TV by first summing the squared difference of pixel-intensities in the horizontal direction and in the vertical direction; followed by normalized by the number of elements ($CHW$). In our experiments, we scale the final TV loss using a hyperparameter $\lambda_{\text{TV}}$, which controls the strength of the regularization.

*4) Color Constancy Loss:* This loss component attempts to enforce consistent color balance across the RGB channels and suppress color distortions. In this way, we expect the means of the red, green, and blue channels to be close to each other in the enhanced image. Given an output image $x \in \mathbb{R}^{N\times3\times H\times W}$, the mean of each channel is computed spatially across height and width. Let $\mu_r, \mu_g, \mu_b$ denote the mean intensities of the red, green, and blue channels, respectively. The loss is defined as follows:

$$\mathcal{L}_{\text{color}} = \lambda_c \cdot \left[(\mu_r - \mu_g)^2 + (\mu_r - \mu_b)^2 + (\mu_g - \mu_b)^2\right] \tag{9}$$

where $\lambda_c$ is a scalar weight to control the influence of this term. The loss penalizes large deviations between color channels, promoting a natural-looking white balance. In our case $\lambda_c$ is kept at 0.5.

*5) Gamma Regularization Loss:* To prevent instability due to extreme or unnatural gamma corrections, we employ a *Gamma Regularization Loss*, which constrains the predicted gamma maps toward a desired average. We first compute the global mean gamma value $\bar{\gamma}$ across all channels and spatial locations. Next, we regularize it toward a predefined target value $\gamma_t$ (in our case, 1.0). The loss is defined as follows:

$$\mathcal{L}_{\gamma} = \lambda_{\gamma} \cdot (\bar{\gamma} - \gamma_t)^2 \tag{10}$$

Here, $\lambda_\gamma$ is a weighting coefficient that controls the influence of regularization. This loss penalizes deviations from the desired global gamma intensity, promoting a perceptually stable and consistent brightness correction across the image.

**Dual-Stage Composite Loss Function:**
Our training objective is governed by a *Dual-Stage Loss Function*, which supervises both the intermediate gamma-corrected output and the final enhanced image. This design enables stage-wise learning while prioritizing the final perceptual quality. Let the outputs of the two enhancement stages be denoted by $x_t$ (gamma-enhanced output). Recall, final enhanced output and the ground-truth images are referred as $\hat{(x)}$ and $x$ respectively. The dual-stage loss incorporates multiple components at each stage: pixel-level reconstruction (L1), structural similarity (SSIM), spatial smoothness (TV), and color constancy. Additionally, a gamma regularization term is applied to the predicted gamma maps in the first stage to stabilize gamma values.

The Stage-1 loss (applied to $x_t$) is defined as:

$$\begin{aligned}\mathcal{L}_{\text{stage1}} = \alpha \cdot \mathcal{L}_{\text{L1}}(x_t, x) + \beta \cdot \mathcal{L}_{\text{SSIM}}(x_t, x) \\ + \gamma \cdot \mathcal{L}_{\text{TV}}(x_t) + \delta \cdot \mathcal{L}_{\text{color}}(x_t) + \mathcal{L}_\gamma(\Gamma),\end{aligned} \tag{11}$$

where $\Gamma$ is the predicted gamma correction map. The Stage-2 loss (applied to $\hat{x}$) is defined similarly, but without the gamma regularization:

$$\begin{aligned}\mathcal{L}_{\text{stage2}} = \alpha \cdot \mathcal{L}_{\text{L1}}(\hat{x}, x) + \beta \cdot \mathcal{L}_{\text{SSIM}}(\hat{x}, x) \\ + \gamma \cdot \mathcal{L}_{\text{TV}}(\hat{x}) + \delta \cdot \mathcal{L}_{\text{color}}(\hat{x})\end{aligned} \tag{12}$$

To prioritize the final stage while still guiding intermediate corrections, the total loss is computed as a weighted sum of both stages:

$$\mathcal{L}_{\text{total}} = 0.3 \cdot \mathcal{L}_{\text{stage1}} + 0.7 \cdot \mathcal{L}_{\text{stage2}}$$

Here, $\alpha, \beta, \gamma, \delta$ are hyper-parameters that control the relative contributions of the loss components. In our implementation, we empirically set $\alpha = 0.5$, $\beta = 0.2$, $\gamma = 0.2$, and $\delta = 0.1$, balancing the trade-off between pixel-wise accuracy, perceptual similarity, spatial smoothness, and color consistency.

## IV. Experiments

### A. Experimental Settings

*1) Dataset:* For evaluation of our GAtED method, we conducted experiments on 3 widely used datasets with paired low-light and normal-light images: LOL-v1 [9] , LOL-v2 real [42] and LOL-v2 synthetic [42]. The LOL-v1 dataset comprises 485 pairs of low-light and normal-light images for training and 15 pairs for testing. Each image has a spatial resolution of 600 × 400 pixels, and all samples were captured under real-world illumination conditions. The LOL-v2 data set is divided into two subsets: LOL-v2 Real and LOL-v2 Synthetic (Syn). The LOL-v2 Real subset contains 689 training pairs and 100 testing pairs of real low-light and corresponding normal-light images. In contrast, the LOL-v2 Syn subset consists of 900 training pairs and 100 testing pairs, where the low-light images are synthetically generated from normal-light counterparts to simulate diverse lighting degradations. To enhance the robustness of the model under varying lighting conditions, we combined both subsets, resulting in a total of 2,074 training pairs and 215 testing pairs in order to compare with various no-reference image quality assessment metrics.

We further validate our model's generalizability by comparing against standard low-light datasets without ground-truth: DICM [43], LIME [8], MEF [44], and NPE [45]. The datasets DICM contains 64 low-light images of mainly landscape scenes, LIME contains 10 images mainly focused on dark street landscapes, MEF has 17 images of dark indoor and building scenes, NPE has 8 low-light images covering mostly natural sceneries.

*2) Metrics:* For evaluation of the model performance we used metrics - PSNR (Peak Signal to Noise Ratio), SSIM (Stuctural Similarity Index Measure) [47], LPIPS (Learned Perceptual Image Patch Similarity) [48], MAE (Mean Absolute Error) Peak Signal-to-Noise Ratio (PSNR) quantifies the relationship between the maximum possible signal power of an image and the power of the noise that affects its fidelity. Structural Similarity Index (SSIM) evaluates the perceptual similarity between the enhanced image and its reference image based on luminance, contrast, and structure. Learned Perceptual Image Patch Similarity (LPIPS) measures perceptual differences between images using deep feature representations. The mean absolute error (MAE) computes the average absolute difference between the enhanced and reference images.

We also employ several no-reference image quality assessment (NR-IQA) metrics, including the Natural Image Quality Evaluator (NIQE) [49], which quantifies image quality based on deviations from natural scene statistics without requiring any training on human-rated data; the Unified No-reference Image Quality and Uncertainty Evaluator (UNIQUE) [50], a deep learning–based model designed to assess image quality across both synthetic and realistic distortion scenarios; and the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [51], which measures perceptual quality degradation by analyzing locally normalized luminance statistics in the spatial domain.

*3) Setup details:* Our method is implemented using Pytorch and NVIDIA-RTX 3050 Ti GPU. During training we have used 100 epochs and all the images are resized to 128 X 128 pixels followed by normalization of the pixels to 0 to 1 range. For the backbone model of LPIPS we have use AlexNet.

### B. Quantitative Comparison with Benchmarking Methods

To thoroughly evaluate the effectiveness of our proposed approach, we conduct comprehensive comparisons with a diverse set of state-of-the-art low-light image enhancement methods across multiple benchmarks and evaluation metrics. The methods includes Retinex based methods LIME [8], RetinexNet [9], KinD [11], RUAS [13], CUE [16], RetinexFormer [15], C-Retinex [14], U-RetinexNet [10], Diff-Retinex++ [17], U-RetinexNet++ [12] followed by diffusion based methods - CLE Diffusion [23], PyDiff [24], GSAD [26], Quad Prior [27], DiffUIR [25], and other image enhancement methods ZeroDCE [21], EnlightenGAN [22], Restormer [29], NeRCo [31], Restormer [29], LLFormer [30],

TABLE I: Quantitative Comparison of Our GAtED and Existing Low-Light Image Enhancement Methods on the LOLv1 [9], LOLv2 Real and LOLv2 Syn Datasets [42].

| **Methods** | **Venue** | **LOLv1 Dataset** | | | | **LOLv2 real Dataset** | | | | **LOLv2 syn Dataset** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PSNR | SSIM | LPIPS | MAE | PSNR | SSIM | LPIPS | MAE | PSNR | SSIM | LPIPS | MAE |
| LIME [8] | TIP'17 | 16.554 | 0.429 | 0.405 | 0.123 | 15.105 | 0.402 | 0.426 | 0.145 | 16.570 | 0.740 | 0.210 | 0.121 |
| RetinexNet [9] | BMVC'18 | 17.558 | 0.651 | 0.379 | 0.117 | 16.097 | 0.401 | 0.543 | 0.131 | 17.137 | 0.762 | 0.255 | 0.117 |
| KinD [11] | ACMMM'19 | 17.648 | 0.775 | 0.175 | 0.123 | 16.828 | 0.759 | 0.224 | 0.129 | 18.930 | 0.809 | 0.174 | 0.107 |
| RUAS [13] | CVPR'21 | 16.405 | 0.500 | 0.270 | 0.153 | 15.326 | 0.488 | 0.310 | 0.162 | 13.404 | 0.645 | 0.364 | 0.196 |
| Zero-DCE [21] | CVPR'20 | 19.524 | 0.703 | 0.330 | 0.106 | 18.059 | 0.574 | 0.313 | 0.131 | 17.756 | 0.816 | 0.168 | 0.124 |
| EnlightenGAN [22] | TIP'21 | 17.483 | 0.651 | 0.322 | 0.135 | 18.640 | 0.676 | 0.309 | 0.132 | 16.573 | 0.775 | 0.212 | 0.133 |
| Restormer [29] | CVPR'22 | 22.156 | 0.817 | 0.151 | 0.078 | 21.236 | 0.820 | 0.191 | 0.082 | 25.105 | 0.925 | 0.066 | 0.057 |
| LLFormer [30] | AAAI'23 | 23.649 | 0.819 | 0.169 | 0.063 | 20.154 | 0.809 | 0.207 | 0.070 | 24.229 | 0.920 | 0.063 | 0.065 |
| Retinexformer [15] | ICCV'23 | 23.386 | 0.833 | 0.140 | 0.067 | 22.254 | 0.831 | 0.112 | 0.074 | 25.488 | 0.930 | 0.061 | 0.054 |
| NeRCo [31] | ICCV'23 | 22.946 | 0.786 | 0.146 | 0.069 | 18.490 | 0.633 | 0.414 | 0.111 | 19.510 | 0.763 | 0.231 | 0.095 |
| CUE [16] | ICCV'23 | 21.670 | 0.775 | 0.224 | 0.079 | 18.053 | 0.753 | 0.347 | 0.129 | 22.209 | 0.877 | 0.117 | 0.072 |
| C-Retinex [14] | IJCV'24 | 19.866 | 0.807 | 0.196 | 0.096 | 18.265 | 0.789 | 0.285 | 0.122 | 20.366 | 0.876 | 0.125 | 0.086 |
| CLE Diffusion [23] | ACMMM'23 | 21.282 | 0.794 | 0.157 | 0.080 | 21.995 | 0.798 | 0.191 | 0.088 | 20.284 | 0.852 | 0.102 | 0.108 |
| QuadPrior [27] | CVPR'24 | 18.771 | 0.781 | 0.209 | 0.112 | 20.471 | 0.811 | 0.198 | 0.081 | 16.102 | 0.758 | 0.251 | 0.146 |
| PyDiff [24] | arXiv'23 | 23.275 | 0.858 | 0.107 | 0.074 | 21.940 | 0.841 | 0.178 | 0.082 | 22.989 | 0.929 | 0.082 | 0.070 |
| GSAD [26] | NeurIPS'23 | 22.978 | 0.851 | 0.103 | 0.076 | 20.190 | 0.847 | 0.112 | 0.102 | 24.219 | 0.927 | 0.052 | 0.065 |
| DiffUIR [25] | CVPR'24 | 22.407 | 0.834 | 0.157 | 0.073 | 19.710 | 0.825 | 0.211 | 0.105 | 19.611 | 0.863 | 0.160 | 0.103 |
| Diff-Retinex++ [17] | TPAMI'25 | 24.667 | 0.867 | 0.101 | 0.062 | **23.413** | 0.872 | 0.134 | **0.064** | 26.058 | 0.944 | **0.041** | 0.054 |
| LL-UNet++ [20] | TCI'24 | 23.005 | 0.8682 | 0.104 | – | – | – | – | – | 18.02 | 0.687 | 0.347 | – |
| RIRO [18] | TCI'24 | 23.01 | 0.897 | **0.072** | – | – | – | – | – | – | – | – | – |
| DLN [19] | TIP'20 | – | – | – | – | 21.946 | 0.807 | – | – | 23.829 | 0.912 | – | – |
| URetinexNet [10] | CVPR'22 | 21.32 | 0.834 | 1.22 | 0.067 | – | – | – | – | – | – | – | – |
| ALL-E+ [46] | TPAMI'25 | 19.56 | 0.773 | 0.209 | – | – | – | – | – | – | – | – | – |
| URetinex-Net++ [12] | TPAMI'25 | 23.826 | 0.8411 | 0.2311 | 0.0589 | – | – | – | – | – | – | – | – |
| **GAtED** | **Proposed** | **25.46** | **0.9140** | 0.088 | **0.0467** | 22.50 | **0.9047** | **0.0763** | 0.0762 | **29.96** | **0.9458** | 0.0515 | **0.0324** |

LL-UNet++ [20], RIRO [18], DLN [19], ALL-E+ [46], ISSR [52], MIRNet [53], SNR [54], LLFlow [55], ReLLIE [56], SCL-LLE [57], SCI [58], PairLIE [59]

*1) With availability of ground-truth images:* We conducted experiments on the LOL dataset [9] [42], which are standard benchmarks in the literature on low-light enhancement. The results are shown in Table I. We have reported four metrics (PSNR, SSIM, LPIPS, and MAE) to quantifying the similarity between the enhanced and ground-truth images. Our method consistently demonstrates competitive or superior performance across most metrics. In particular, our method GAtED surpasses all the metrics except LPIPS for LOL-V1 dataset, PSNR and MAE for LOL-v2 real dataset and LPIPS for LOL-v2 Synthetic dataset. It is important to note that our method achieves remarkably high SSIM values for all variants of LOL dataset.

*2) Without availability of ground-truth images:* Furthermore, to assess perceptual quality in the absence of ground truth, evaluations were conducted on standard no-reference low-light datasets, including DICM [43], LIME [8], MEF [44], and NPE [45].. The performance was quantified using NIQE, BRISQUE, and UNIQUE metrics, which provide a comprehensive measure of visual quality. As shown in Table II, GAtED achieves the best NIQE scores across all datasets, indicating that it produces images with perceptual characteristics most closely aligned with natural scene statistics. Our outstanding NIQE performance across no-reference datasets (DICM: 2.96, LIME: 3.76, MEF: 2.76, NPE: 3.09) demonstrates the natural appearance and artifact-free quality of our enhanced images. This superior NIQE performance reflects the model's ability to effectively enhance illumination while preserving structural fidelity and reducing unnatural artifacts, thereby yielding visually pleasing and realistic results.

Overall, our proposed model achieves consistently strong performance across all evaluated datasets (both types: with and without ground-truth images) and metrics. This demonstrates its capability to generalize well to varying lighting conditions and its effectiveness in both pixel-based and perceptual quality enhancement.

### C. Visual Comparision

To demonstrate the effectiveness of our proposed method, we visually compared it with some of the state-of-the-art methods, KinD [11], ZeroDCE [21], RetinexNet [9], U-RetinexNet [10], NeRCo [31], in various challenging scenarios. We present visual comparisons on two low-light images (with varying low-lightness) in Figures 3 and 4.

The night scene (input image) in Figure 3 is captured under extreme low-lighting conditions. Although it is hardly visible in the low-light image, the ground-truth image confirms the presence of fine details in the green grass region. The output images of most of the methods compared failed to successfully preserve fine structures and textures without color distortion. Unlike RetinexNet which tends to produce overexposed regions in brightly lit areas, or KinD which often results in underexposed dark regions, our method GAtED provides content-aware illumination adjustment that preserves the natural luminance distribution across different image regions.

In Figure 4, we display the results on an image captured in the early morning. Although it has a relatively better lighting condition than the previous low-light image, this image is rich in strong structures. With close visual comparison with reference to the ground-truth, it is evident that most of the existing methods (shown here) suffer from color distortion.

TABLE II: Quantitative comparison of GAtED and denoising results on DICM, LIME, MEF, and NPE datasets, including NIQE↓, BRISQUE (BR.)↓, and UNIQUE (UN.)↑.

| Methods | Venue | DICM [43] | | | LIME [8] | | | MEF [44] | | | NPE [45] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NIQE | BR. | UN. | NIQE | BR. | UN. | NIQE | BR. | UN. | NIQE | BR. | UN. |
| LIME | TIP'17 | 3.75 | 24.99 | 0.78 | 3.85 | 18.65 | 0.53 | 3.65 | 18.10 | 0.65 | 4.44 | 17.73 | 0.93 |
| Retinex-Net | BMVC'18 | 4.47 | 30.38 | 0.75 | 4.60 | 26.42 | 0.52 | 4.41 | 21.90 | 0.97 | 4.60 | 22.59 | 0.81 |
| ISSR | ACMMM'20 | 4.14 | 26.93 | 0.59 | 4.17 | 18.82 | 0.83 | 4.22 | 25.84 | 0.87 | 4.02 | 21.27 | 0.99 |
| MIRNet | ECCV'20 | 5.15 | 37.56 | -0.19 | 5.59 | 37.93 | 0.31 | 4.38 | 33.81 | 0.48 | 5.41 | 31.05 | 0.16 |
| URetinex-Net | CVPR'22 | 3.95 | 25.24 | 0.85 | 4.34 | 26.48 | 0.93 | 3.79 | 21.08 | 1.18 | 4.69 | 27.84 | 0.99 |
| SNR | CVPR'22 | 4.62 | 31.05 | 0.12 | 4.49 | 31.65 | 0.53 | 4.09 | 27.18 | 0.53 | 4.36 | 22.59 | 0.36 |
| LLFlow | AAAI'22 | 3.73 | 24.29 | **1.02** | 3.82 | 26.53 | **1.03** | 3.84 | 25.67 | **1.39** | 4.23 | 16.78 | **1.28** |
| Zero-DCE | CVPR'20 | 3.56 | 25.52 | 0.82 | 3.77 | 21.10 | 0.73 | 3.28 | 19.83 | 1.22 | 3.95 | 19.28 | 1.07 |
| RUAS | CVPR'21 | 5.21 | 32.58 | -0.17 | 4.26 | 22.35 | 0.34 | 3.83 | 18.82 | 0.73 | 5.14 | 22.85 | 0.37 |
| ReLLIE | ACMMM'21 | 4.44 | 27.00 | 0.41 | 5.22 | 20.27 | 0.52 | 5.22 | 27.74 | 1.07 | 4.47 | 28.68 | 0.21 |
| SCL-LLE | AAAI'22 | 3.51 | 24.63 | 0.87 | 4.01 | 16.99 | 0.76 | 3.31 | 16.18 | 1.25 | 4.47 | 20.51 | 0.38 |
| SCI | CVPR'22 | 4.11 | 26.83 | 0.39 | 4.21 | **14.78** | 0.62 | 4.09 | 14.76 | 0.65 | 4.28 | 25.93 | 0.94 |
| PairLIE | CVPR'23 | 4.08 | 31.80 | 0.67 | 4.52 | 22.59 | 0.78 | 4.10 | 29.56 | 1.08 | 4.19 | 23.12 | -0.01 |
| NeRCo | ICCV'23 | 4.37 | 32.24 | 0.04 | 4.86 | 22.65 | 0.16 | 4.06 | 30.70 | 1.06 | 4.14 | 15.19 | 0.96 |
| ALL-E | TPAMI | 3.49 | 24.56 | 0.88 | 3.78 | 21.48 | 0.80 | 3.32 | 30.46 | 1.27 | 3.85 | 18.58 | 1.10 |
| ALL-E+NIMA | TPAMI | 3.47 | 24.99 | 0.87 | 3.77 | 21.42 | 0.69 | 3.32 | 30.27 | 1.28 | 4.16 | **14.45** | 0.94 |
| ALL-E+TANet | TPAMI | 3.81 | **21.43** | 0.87 | 3.84 | 18.52 | 0.61 | 3.26 | **13.16** | 1.29 | 4.14 | 15.19 | 0.96 |
| **GAtED** | Proposed | **2.96** | 25.49 | 0.771 | **3.76** | 22.47 | 0.82 | **2.76** | 23.58 | 0.76 | **3.09** | 18.36 | 1.01 |

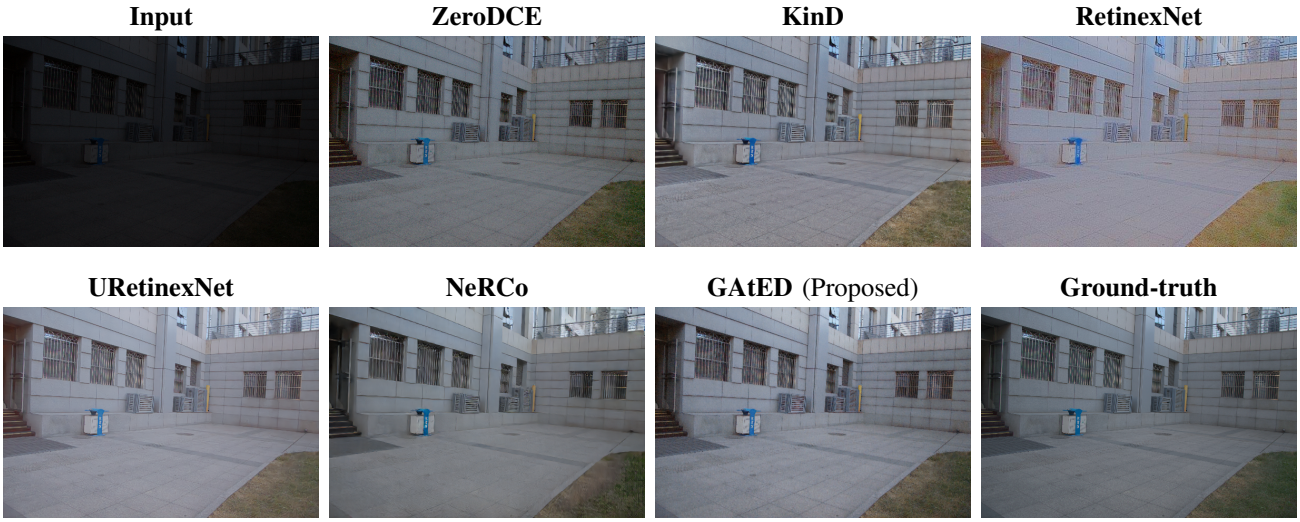


Fig. 3: Visual comparison of different low-light image enhancement methods under extreme low light condition. Notice the appearance of unwanted color distortion in the output images of most of the methods compared. Unlike RetinexNet which tends to produce overexposed regions in brightly lit areas, our method GAtED produced content-aware illumination adjustment that preserves the natural luminance distribution across different image regions.

In particular, there is a prominent lack of brightness in the outputs of ZeroDCE and NeRCo, whereas KinD also has a slight drop in brightness. The outputs of both RetinexNet and URetinexNet exhibit over-enhancement. In fact, RetinexNet results in significant structural distortion. In contrast, our GAtED method consistently achieves a superior brightness enhancement while maintaining a natural appearance.

In summary, both visual evaluations encompass both the improvement of overall image quality and detailed preservation of fine structures and textures. A critical advantage of our method is the preservation of color consistency and the prevention of color distortion artifacts. As shown in the comparison images, methods like RetinexNet and NeRCo exhibit noticeable color shifts, particularly in regions with complex lighting conditions. Our attention-enabled encoder decoder network effectively maintains color balance through its attention-guided refinement process, resulting in enhanced images that closely match the ground truth color characteristics. The visual comparisons on these datasets show that our method produces images with characteristics most closely aligned with natural scene statistics, avoiding the over-enhancement and artificial appearance that plague other methods

## V. Conclusion

In this work, we proposed a novel two-stage deep learning method for low-light image enhancement. that effectively inte-

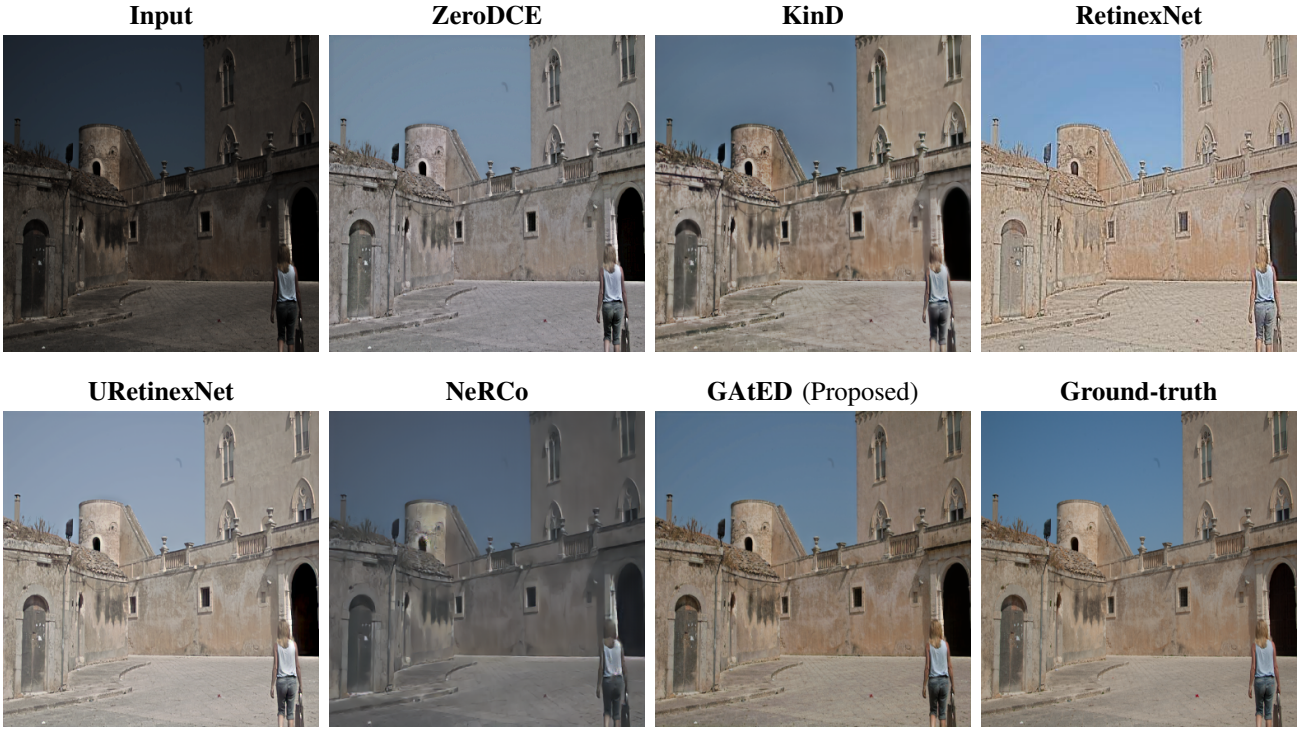


Fig. 4: Visual comparison of different low-light image enhancement methods under moderate darkness. It is visually evident that our GAtED method consistently achieved a superior brightness enhancement while maintaining a natural appearance.

grates adaptive illumination correction and attention-guided refinement. The core idea was to first perform a learnable gamma correction which was followed by incorporating an attention mechanism in an encoder-decoder deep network. The first deep stage adaptively learned content-aware gamma maps to recover overall visibility, while the attention-enabled encoder-decoder network recovered structural detail, suppressed noise, and corrected color imbalances. Experimental results confirmed that our model could address both global brightness inconsistencies and local perceptual degradations. It consistently achieved competitive performance on multiple image quality metrics (PSNR, SSIM, LPIPS, MAE, NIQE, BRISQUE, and UNIQUE) compared to existing methods.

In future work, we plan to extend the proposed method for enhancing low-light videos. The process of video capturing in real often leads to blurring (for example, hand-shaking blur, motion blur) of the video frames. Therefore, we aim to solve the joint deblurring and enhancement of low-light image and video. We also plan to translate our method into practical applications, such as mobile phones and embedded devices, to make it suited to a wide range of applications such as surveillance and autonomous vehicles.

## REFERENCES

[1] X. Xu, S. Wang, Z. Wang, X. Zhang, and R. Hu, "Exploring image enhancement for salient object detection in low light images," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 17, no. 1s, pp. 1–19, 2021.

[2] C. Li, X. Qu, A. Gnanasambandam, O. A. Elgendy, J. Ma, and S. H. Chan, "Photon-limited object detection using non-local feature matching and knowledge distillation," *Proc. IEEE/CVF International Conference on Computer Vision*, pp. 3976–3987, 2021.

[3] J. Liu, D. Xu, W. Yang, M. Fan, and H. Huang, "Benchmarking low-light image enhancement and beyond," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1153–1184, 2021.

[4] S.-D. Chen and A. R. Ramli, "Minimum mean brightness error bi-histogram equalization in contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 4, pp. 1310–1319, 2003.

[5] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1752–1758, 2007.

[6] S. Ghosh, R. G. Gavaskar, D. Panda, and K. N. Chaudhury, "Fast scale-adaptive bilateral texture smoothing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 2015–2026, 2019.

[7] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2782–2790, 2016.

[8] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2016.

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

[10] 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," *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5901–5910, 2022.

[11] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," *Proc. 27th ACM International Conference on Multimedia*, pp. 1632–1640, 2019.

[12] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "Interpretable optimization-inspired unfolding network for low-light image enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 4, pp. 2545 – 2562, 2025.

[13] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement," *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10 561–10 570, 2021.

[14] H. Xu, H. Zhang, X. Yi, and J. Ma, "CRetinex: A progressive color-shift aware retinex model for low-light image enhancement," *International Journal of Computer Vision*, vol. 132, no. 9, pp. 3610–3632, 2024.

[15] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," *Proc. IEEE/CVF International Conference on Computer Vision*, pp. 12 504–12 513, 2023.

[16] N. Zheng, M. Zhou, Y. Dong, X. Rui, J. Huang, C. Li, and F. Zhao, "Empowering low-light image enhancer through customized learnable priors," *Proc. IEEE/CVF International Conference on Computer Vision*, pp. 12 559–12 569, 2023.

[17] X. Yi, H. Xu, H. Zhang, L. Tang, and J. Ma, "Diff-Retinex++: Retinex-driven reinforced diffusion model for low-light image enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[18] L. Zhao, B. Chen, J. Zhang, A. Wang, and H. Bai, "RIRO: From retinex-inspired reconstruction optimization model to deep low-light image enhancement unfolding network," *IEEE Transactions on Computational Imaging*, vol. 10, pp. 969–983, 2024.

[19] L.-W. Wang, Z.-S. Liu, W.-C. Siu, and D. P. Lun, "Lightening network for low-light image enhancement," *IEEE Transactions on Image Processing*, vol. 29, pp. 7984–7996, 2020.

[20] P. Shi, X. Xu, X. Fan, X. Yang, and Y. Xin, "LL-UNet++: Unet++ based nested skip connections network for low-light image enhancement," *IEEE Transactions on Computational Imaging*, vol. 10, pp. 510–521, 2024.

[21] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1780–1789, 2020.

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

[23] Y. Yin, D. Xu, C. Tan, P. Liu, Y. Zhao, and Y. Wei, "CLE diffusion: Controllable light enhancement diffusion model," *Proc. 31st ACM International Conference on Multimedia*, pp. 8145–8156, 2023.

[24] D. Zhou, Z. Yang, and Y. Yang, "Pyramid diffusion models for low-light image enhancement," *arXiv preprint arXiv:2305.10028*, 2023.

[25] D. Zheng, X.-M. Wu, S. Yang, J. Zhang, J.-F. Hu, and W.-S. Zheng, "Selective hourglass mapping for universal image restoration based on diffusion model," *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25 445–25 455, 2024.

[26] J. Hou, Z. Zhu, J. Hou, H. Liu, H. Zeng, and H. Yuan, "Global structure-aware diffusion process for low-light image enhancement," *Advances in Neural Information Processing Systems*, vol. 36, pp. 79 734–79 747, 2023.

[27] W. Wang, H. Yang, J. Fu, and J. Liu, "Zero-reference low-light enhancement via physical quadruple priors," *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26 057–26 066, 2024.

[28] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," *Proc. European Conference on Computer Vision (ECCV)*, pp. 3–19, 2018.

[29] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5728–5739, 2022.

[30] T. Wang, K. Zhang, T. Shen, W. Luo, B. Stenger, and T. Lu, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," *Proc. AAAI Conference on Artificial Intelligence*, vol. 37, no. 3, pp. 2654–2662, 2023.

[31] S. Yang, M. Ding, Y. Wu, Z. Li, and J. Zhang, "Implicit neural representation for cooperative low-light image enhancement," *Proc. IEEE/CVF International Conference on Computer Vision*, pp. 12 918–12 927, 2023.

[32] J. Yang, Y. Xu, H. Yue, Z. Jiang, and K. Li, "Low-light image enhancement based on retinex decomposition and adaptive gamma correction," *IET Image Processing*, vol. 15, no. 5, pp. 1189–1202, 2021.

[33] Y. Wang, Z. Liu, J. Liu, S. Xu, and S. Liu, "Low-light image enhancement with illumination-aware gamma correction and complete image modelling network," *Proc. IEEE/CVF International Conference on Computer Vision*, pp. 13 128–13 137, 2023.

[34] 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.

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

[36] D. J. Jobson, Z.-u. 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.

[37] M. Elad, "Retinex by two bilateral filters," *Proc. International Conference on Scale-Space Theories in Computer Vision*, pp. 217–229, 2005.

[38] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 1997.

[39] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo, "Modeling visual attention via selective tuning," *Artificial Intelligence*, vol. 78, no. 1-2, pp. 507–545, 1995.

[40] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[41] F. Lv, Y. Li, and F. Lu, "Attention guided low-light image enhancement with a large scale low-light simulation dataset," *International Journal of Computer Vision*, vol. 129, no. 7, pp. 2175–2193, 2021.

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

[43] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2D histograms," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5372–5384, 2013.

[44] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3345–3356, 2015.

[45] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.

[46] D. Liang, Y. Gao, L. Li, Z. Xu, S.-J. Huang, and S. Chen, "Aesthetics-guided low-light enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 7, pp. 5866 – 5883, 2025.

[47] 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, no. 4, pp. 600–612, 2004.

[48] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.

[49] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[50] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Transactions on Image Processing*, vol. 30, pp. 3474–3486, 2021.

[51] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[52] M. Fan, W. Wang, W. Yang, and J. Liu, "Integrating semantic segmentation and retinex model for low-light image enhancement," *Proc. 28th ACM International Conference on Multimedia*, pp. 2317–2325, 2020.

[53] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," *Proc. European Conference on Computer Vision*, pp. 492–511, 2020.

[54] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "SNR-aware low-light image enhancement," *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17 714–17 724, 2022.

[55] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. Kot, "Low-light image enhancement with normalizing flow," *Proc. AAAI Conference on Artificial Intelligence*, vol. 36, no. 3, pp. 2604–2612, 2022.

[56] R. Zhang, L. Guo, S. Huang, and B. Wen, "ReLLIE: Deep reinforcement learning for customized low-light image enhancement," *Proc. 29th ACM International Conference on Multimedia*, pp. 2429–2437, 2021.

[57] D. Liang, L. Li, M. Wei, S. Yang, L. Zhang, W. Yang, Y. Du, and H. Zhou, "Semantically contrastive learning for low-light image enhancement," *Proc. AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, pp. 1555–1563, 2022.

[58] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5637–5646, 2022.

[59] Z. Fu, Y. Yang, X. Tu, Y. Huang, X. Ding, and K.-K. Ma, "Learning a simple low-light image enhancer from paired low-light instances," *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22 252–22 261, 2023.