International Journal of Computers and Informatics

Journal Homepage: https://www.ijci.zu.edu.eg

Int. j. Comp. Info. Vol. 9 (2025) 106–122

Paper Type: Original Article

DIAA: Distillation with Illumination-aware Adaptive Attention for Low-Light Image Enhancement

Mostafa Khalifa [1,*], Nabil A. Lashin [1], Hanaa M. Hamza [1] and Khalid M. Hosny [1]

[1] Department of Information Technology, Faculty of Computer and Informatics, Zagazig University, Zagazig 44519, Egypt.
Emails: makhalifa@fci.zu.edu.eg; nlashin@yahoo.com; hanaa_hamza2000@yahoo.com; k_hosny@zu.edu.eg.

Received: 01 Sep 2025 **Revised:** 02 Oct 2025 **Accepted:** 08 Dec 2025 **Published:** 11 Dec 2025

Abstract

Low-light image enhancement plays a vital role in improving images obtained under poor illumination conditions encountered in photography, surveillance, and autonomous systems. Such images often exhibit low contrast, amplified noise, and color distortion, which adversely affect downstream tasks. In this paper, we transfer knowledge from a high-performing teacher network to a smaller student network designed for low-light image enhancement. This student model is a hybrid deep-learning architecture integrating adaptive exposure correction, trainable gamma modulation, multi-scale feature extraction, spatial and frequency attention, noise-aware residual blocks, and transformer-CNN fusion. A key innovation in our paper is the use of a hybrid loss function that combines pixel-level accuracy, perceptual quality, structural consistency, and illumination stability. It includes MSE, L1, SSIM, VGG19-based perceptual loss, illumination smoothness, color constancy, and exposure-control terms. Furthermore, when a teacher model is available, a temperature-scaled knowledge distillation loss transfers soft supervision to the student network. This multi-term objective enables the model to restore natural brightness, preserve details, maintain color balance, and suppress noise. Comprehensive experiments on benchmark datasets demonstrate that our method outperforms state-of-the-art techniques in terms of PSNR, SSIM, and perceptual quality.

Keywords: Low-light; Image Enhancement; Knowledge Distillation; Deep Learning; Adaptive Exposure Correction.

1 | Introduction

Imaging under low-light conditions is challenging due to insufficient illumination, leading to low visibility, poor contrast, noise amplification, and color distortions. Such issues reduce image quality and hinder the effectiveness of applications such as video surveillance, autonomous driving, and photography [1].

Low-light image enhancement (LLIE) seeks to restore image visibility under difficult lighting conditions, and the process is even more challenging than traditional noise reduction. In most noise-reduction scenarios, additive or multiplicative noise can be reduced by estimating and reversing the underlying degradation process. For LLIE, instead of a single noise source, the task involves simultaneous correction of lighting, contrast, color fidelity, and noise, making the restoration challenge much more complex.

In low-light real-world conditions, it is likely that combined degradations from photon-dominated noise, low demodulated signal amplitude due to the limited dynamic-range sensitivity of sensors, and intensity-

Corresponding Author: makhalifa@fci.zu.edu.eg

dependent noise will occur. As a result, simple brightness amplification tends to enhance noise visibility, distort color relationships, and reduce texture fidelity. This is an important area of research, and the development of robust LLIE algorithms is an active research area. Effective enhancement must balance lighting correction, noise reduction, and detail retention while performing consistently across a wide range of exposure conditions [2]. Our work expands on this goal by seeking to achieve consistent increases in both visual quality and downstream task performance.

As depicted in Figure 1 Several low-light image enhancement approaches have been developed to improve image contrast and brightness. They can be classified into two main categories: traditional enhancement methods and learning-based methods. There are three types of conventional approaches: histogram-based methods [3–6], gamma correction-based methods [7–9], and Retinex-based methods [10, 11]. Learning-based methods are further divided into Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs).

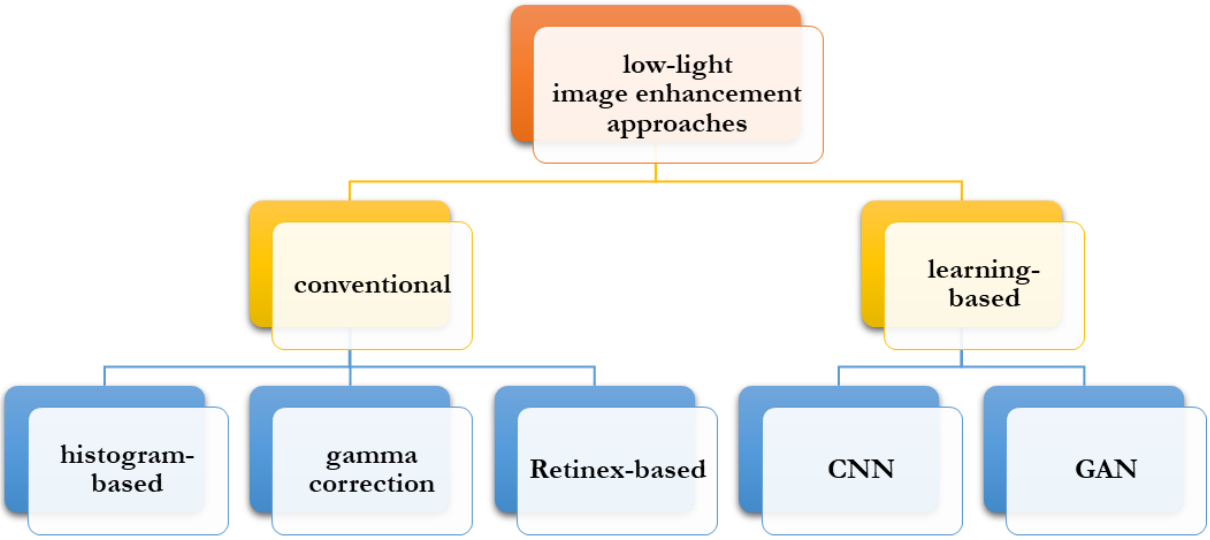

Figure 1. Low-light image enhancement methods.

Conventional enhancement techniques include histogram equalization-based methods, simple gamma correction-based methods, and Retinex-based methods. The first approach that focuses on low-light image enhancement is Histogram Equalization (HE) [12]. The methods reassign pixel values to achieve a uniform pixel intensity distribution, which typically results in a significant loss of image detail. Gamma correction-based methods enhance dynamic range by applying an exponential transformation to each pixel, improving contrast and brightness. However, treating pixels in isolation without considering surrounding pixels makes these techniques more susceptible to various artifacts and noise amplification. Retinex-based methods separate an image into illumination and reflectance layers and enhance them individually. The fixed constraints used in these methods often lack the flexibility needed for real-world applications, which can lead to color distortion and amplified noise and produce either over-enhanced regions or insufficiently enhanced regions [13].

In recent years, learning- based methods have improved accuracy, speed, and effectiveness in low-light enhancement tasks compared to conventional methods. Most of these innovative strategies predominantly leverage the capabilities of convolutional neural networks or generative adversarial networks, which have been proven to be particularly effective in this context. Furthermore, these advanced techniques can be integrated with the Retinex model and subsequently enhanced using either a recurrent neural network or an encoder-decoder architecture, thereby facilitating substantial improvements in the quality of images captured under low-light conditions. Nevertheless, it is imperative to acknowledge that these sophisticated algorithms require

access to large-scale datasets for optimal performance and, because of their operational intricacies, often result in images that are excessively bright or overexposed.

This work proposes novel hybrid architecture leveraging adaptive exposure correction, trainable gamma correction, noise-aware residual blocks, multi-scale convolutions, spatial and frequency attention modules, and transformer-CNN fusion. The model is trained using a composite loss function that balances pixel accuracy, structural similarity, perceptual quality, and noise suppression. Our experiments show consistent improvements over prior works in both objective metrics and visual quality. Our study's main contributions are as follows :

- Intelligent Hybrid Architecture for Low-Light Image Enhancement: In this manuscript, we present an intelligent hybrid deep network comprising novel trainable preprocessing layers (Gamma Correction Layer and Adaptive Exposure Correction Layer), multi-scale feature extraction, noise modeling via attention residual blocks, frequency-driven feature segmentation, and transformers-convolutional neural network fusion for global and local context modeling for low-light images .
- Personalized Attention Modules: We also use carefully crafted channel and spatial attention structures, together with squeeze-and-excite layers, to control dynamic channel and spatial attention features to enhance important image features critical for successful image brightening correction and denoising.
- Frequency-Aware and Multi-Scale Feature Decomposition: This feature extraction framework also effectively separates illumination features in the frequency domain from other features, including image noise, to enable distinct processing strategies for global image context and complex image details, which contribute to improved brightening correction and image denoising .
- Residual Learning with Trainable Bounds and Activations: The network also uses residual blocks with swish activation functions within trainable bounds to ensure stable network dynamics for efficient gradient flow, allowing modeling using complex network design without reducing image processing efficiency .
- Learnable Residual Brightening Correcting mapping: This approach learns the residual correction mapping uniformly applied to input images for delicate image brightening correction and denoising to retain the original image distribution .
- Resilient Hybrid Loss Function: This network uses a complex loss function to combine various terms, including per-pixel loss terms, MAE loss, L1 loss, and VGG19 perceptual similarity loss, structural similarity loss, illumination smoothness, color correction, and control loss terms via complex functions that can incorporate additional temperature scaled distillation losses for optimal student network training via teacher guidance .
- End-to-End Trainable Hybrid Architecture: This approach also enables each network layer to function as or to work on ‘self-serialized’ Keras layers for seamless integration, simplicity, and simplicity for deployment operations.

2 | Related Work

Low-light image enhancement methods can be broadly categorized into conventional methods and deep learning-based methods.

2.1 | Conventional Methods

2.1.1 | Histogram-based Methods

Conventional methods primarily address poor visibility and low contrast by transforming pixel intensities and manipulating local or global histograms. Early works used classical HE to uniformly redistribute intensity values across the available dynamic range, thereby improving overall visual contrast and brightening dark regions. HE is simple; however, it often results in over-enhancement and loss of image details, limiting its direct applicability in sensitive low-light enhancement tasks. To mitigate these drawbacks, several HE variants have been developed. Bi-Histogram Equalization (BBHE) [14]divides the histogram into two sub-histograms based on the mean intensity and equalizes each sub-histogram separately to preserve image brightness and reduce distortion. Building on this concept, Minimum Mean Brightness Error Bi-Histogram Equalization (MMBEBHE) [15] selects the partition point that minimizes brightness deviation from the original image, effectively preserving the original brightness while improving contrast.

Recursive and multi-histogram methods, such as Recursive Sub-Image Histogram Equalization (RSIHE) [16] and Recursively Separated and Weighted Histogram Equalization (RSWHE) [17] segment the histogram into multiple sub-histograms recursively, performing independent equalization to enhance contrast adaptively while preserving local brightness and details. These allow scalable brightness management and reduce undesirable artifacts associated with single-step HE. However, these methods have several limitations. They tend to be recursively intensive due to recursive operations and multiple histogram calculations, limiting real-time applicability. Adaptive Histogram Equalization (AHE)[18] improves upon global methods by dividing the image into contextual regions and performing localized equalization to enhance local contrast, though AHE can amplify noise. Contrast-Limited AHE (CLAHE) [19] addresses AHE's noise amplification problem by limiting contrast enhancement in homogeneous areas, preventing over-saturation, and preserving natural textures. Various modifications to AHE and CLAHE adapt block sizes, clipping limits, and interpolation strategies to enhance local contrast without excessively amplifying noise. However, both methods introduce artifacts and are computationally intensive for large images.

Dynamic histogram equalization (DHE) [20] partitions histograms into sub-blocks based on smoothed local minima, equalizing sub-histograms to avoid washed-out images and maintain detail integrity. Brightness Preserving Dynamic Histogram Equalization (BPDHE) [21] is a block-based variant of DHE that incorporates brightness-preservation constraints within the dynamic equalization framework to improve realism. While BPDHE preserves brightness consistency between the original and enhanced images, it relies heavily on the block size, leading to uneven brightness. Both methods perform poorly on real low-light photos, often amplifying noise, introducing artifacts, and distorting colors, especially in very dark scenes.

2.1.2 | Gamma Correction-based Methods

Gamma correction (GC) is a transformation that expands an image's dynamic range by applying an exponential function to each pixel. Using a technique involving an active contour model and adaptive gamma correction, Ozturk et al. . [9] develop a low-light enhancement algorithm that creates low-light enhanced images by estimating illumination regions within the image through active contour modeling, followed by adjusting brightness non-linearly via adaptive gamma correction across multiple gamma-corrected versions of the image, so the results can be fused, generating images with improved contrast and a more natural look. The proposed algorithm excels with images having relatively low light levels. Still, it is limited by its reliance on handcrafted illumination models and does not provide robust noise performance in naturally occurring, extremely low-light, or noisy environments. Huang et al. [7] introduced an Adaptive Gamma Corrector that uses a weighting distribution function to reshape the histogram and thus avoid over-enhancement, producing significantly better global contrast and more natural-looking images than those made using standard (fixed) gamma corrections; however, it still lacks sufficient robustness to address typical problems associated with images that contain a significant level of non-uniform illumination and high levels of noise. Li et al. [8]

proposed a gamma correction approach that utilizes a pixel-level framework; hence, each pixel's corresponding gamma value is uniquely defined based upon the local context of the pixel and enables fine-grained and content-aware enhancement. Compared to conventional gamma techniques, it retains superior detail and uniformity. Still, it is more susceptible to noise and can struggle in very low-light conditions unless assisted by specially designed noise-reducing technologies.

The main disadvantage of gamma-based techniques is that they do not account for the interrelationships between neighboring pixels or for how these relationships differ across channels, making them inherently more prone to artifacts and noise amplification.

2.1.3 | Retinex-based Methods

Retinex-based algorithms separate images into illumination and reflectance components, gradually enhancing illumination while preserving reflectance to maintain structural consistency and realistic colors [22].

The early foundations of this line of work were established by Jobson et al. through the Retinex theory and the development of the Multi-Scale Retinex (MSR) algorithm [23], followed by the MSRCR variant [24], which introduced a color restoration mechanism inspired by human visual adaptation. These contributions aimed to achieve a balance between dynamic range compression and perceptual color fidelity.

Further research aimed at refining illumination estimation and managing spatial non-uniformity. Wang et al. [25] proposed a method to improve low-light images by estimating a smooth illumination map, enhancing contrast while maintaining natural brightness and color. It effectively avoids halos and performs well in uneven lighting, but faces challenges with computational costs, very dark, noisy regions, and abrupt illumination changes. Fu et al. [26] developed a weighted variational model for improved reflectance and illumination accuracy, enhancing detail retention in darker areas. Guo et al. [27] proposed LIME with channel maxima for pixel-wise illumination estimation with low computation cost.

Nonetheless, traditional Retinex image enhancement techniques rely on handcrafted priors and accurate illumination estimates, which tend to fail in real-world scenarios characterized by noise, color shifts, and compressed histograms. The outcome could lead to problematic consequences, such as color distortion, noise amplification, or improper image enhancement across different areas of different input images, when only using analytical knowledge priors.

2.2 | Learning-Based Methods

Recent low-light enhancement methods cover a broad spectrum, from classical Retinex formulations to modern CNNs, transformers, and flow or diffusion models.

Retinex-based methods (RetinexNet [28] , KinD [29], and KinD++ [30]) separate images into lighting and reflectance components, then enhance the lighting. They work well but are heavy and multistaged, making them impractical for resource-limited devices. Lightweight Curve Methods (Zero-DCE [31], SCI [32], RUAS [33]) learn pixel-wise enhancement curves, which are extremely fast and efficient. However, their hand-crafted rules struggle with severe noise and complex textures. High-capacity GANs such as EnlightenGAN [34] utilize adversarial training for realistic results but suffer from instability during optimization. Despite its compactness, DIAA adds noise-aware processing, frequency fusion, and teacher-guided learning for better results. Modern Approaches such as LLFlow [35], and Retinexformer [36] Flow-based models offer probabilistic mappings but are complex. Transformers capture long-range dependencies for better color consistency but are computationally expensive.

Overall, existing methods trade off interpretability (Retinex-based models), efficiency (Zero-DCE, SCI, RUAS), perceptual realism (GANs, flow/diffusion models), and reconstruction accuracy (modern CNNs, transformers). DIAA builds on these lines of work by combining explicit illumination and frequency modeling with a compact, distillation-driven architecture, aiming to preserve much of the quality of heavier models while remaining suitable for deployment on constrained platforms.

3 | Proposed Method

Knowledge distillation provides the conceptual backbone of our framework. Instead of training a single network in isolation, we explicitly transfer knowledge from a strong but computationally heavy "teacher" to a compact "student" model. In our design, the teacher is a ResNet50 network pretrained on ImageNet and fine-tuned for low-light image enhancement, leveraging its deep residual structure to learn rich, high-level representations of illumination, texture, and scene context.

The student aligns with the proposed DIAA network architecture, which is intentionally lighter and tailored for enhancement via illumination-aware preprocessing, multiscale and frequency-aware feature extraction, and noise-aware residual blocks. During training, the student is supervised not only by ground-truth normal-light images but also by soft predictions from the ResNet50 teacher, trained with a temperature-scaled distillation loss. This setup allows the student to mimic the teacher's nuanced responses while maintaining a substantially lower computational cost, making the resulting model more suitable for real-time or resource-constrained deployments.

The DIAA network, shown in Figure 2, is a fully convolutional model for low-light image enhancement, serving as a compact, computationally efficient alternative to the teacher model. Given a low-light RGB image, the network produces an enhanced image with the exact resolution. The architecture consists of seven main stages: (1) image preprocessing, (2) illumination estimation, (3) multiscale frequency analysis, (4) dual branch feature extraction, (5) noise-aware residual refinement, (6) global-local frequency fusion, and (7) residual enhancement and reconstruction. All components are differentiable and trained jointly with a metric-guided hybrid loss and an optional distillation loss term.

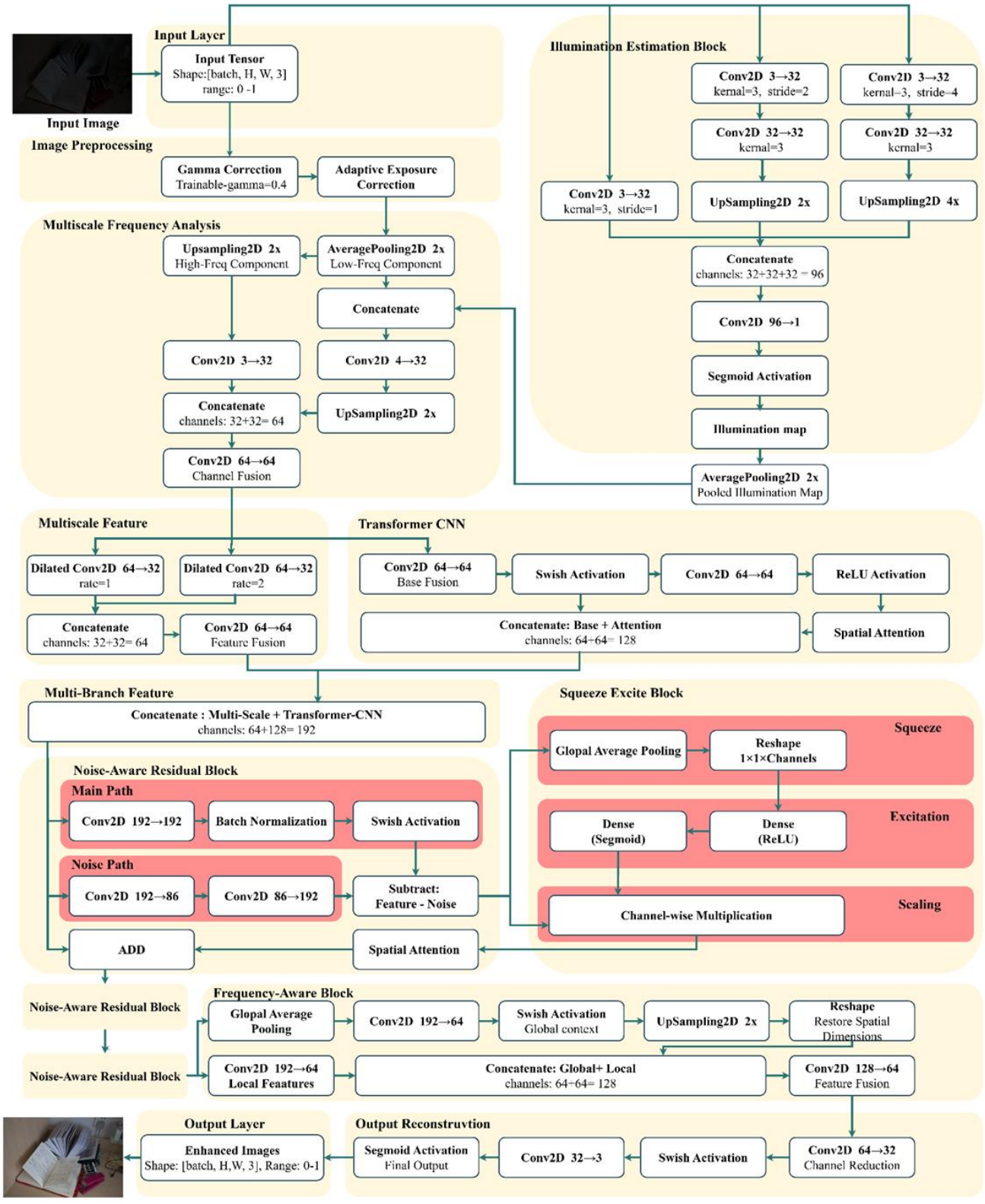

Figure 2. The Structure of the DIAA Network.

3.1 | Image Preprocessing

Trainable Gamma Correction. The first preprocessing step applies a learnable gamma correction to normalize the global tone of the input:

$$I_\gamma = clip(I_{in}, 0,1)^\gamma \quad (1)$$

Here I_{in} is the input image, I_γ is the gamma-corrected image, and γ is a trainable scalar weight initialized to 0.4 and constrained to $[0.1,2.0]$ using a clipping constraint. This parameter allows the network to automatically adjust the nonlinear tone curve rather than relying on fixed gamma values.

Adaptive Exposure Correction. To further regularize global brightness, an adaptive exposure layer multiplies the gamma-corrected image by an exponential gain:

$$I_p = I_\gamma \cdot exp(g) \quad (2)$$

Here g is a learnable scalar initialized to zero. The exponential form guarantees positive gains and enables smooth exposure adjustment. The resulting image I_p serves as the shared input for the illumination estimation and frequency analysis blocks.

3.2 | Illumination Estimation

The illumination estimation branch explicitly models spatially varying lighting. Given I_p, three convolutional paths operate at different scales:

- A full-resolution path with stride 1,
- A half-resolution path with a stride of 2,
- A quarter-resolution path with a stride of 4.

Each path uses 3×3 convolutions with Swish or ReLU activation to extract scale-specific illumination features. The low-resolution features are refined with additional convolutions and upsampled back to the original size using bilinear upsampling. All three feature maps are concatenated along the channel dimension and passed through a 3×3 convolution followed by a sigmoid activation to produce a single-channel illumination map $L \in [0,1]^{H \times W \times 1}$. This illumination map captures scene-level lighting structure and is later pooled and injected into the low-frequency branch, providing a strong prior for illumination-aware enhancement.

3.3 | Multiscale Frequency Analysis

To decouple structure and detail, DIAA performs an explicit decomposition into low- and high-frequency components.

Low-frequency Component. The 2×2 average pooling layer downsamples I_p by a factor of two, producing a coarse estimate of scene structure. In parallel, the illumination map L is average-pooled with the same factor. Finally, they are concatenated:

$$F_{low} = \phi\left(concat\left(AvgPool(I_p), AvgPool(L)\right)\right) \quad (3)$$

Here $\phi(\cdot)$ denotes a 3×3 convolution with Swish activation producing 64 channels.

High-frequency Component. The low-frequency representation is upsampled back to the original resolution and subtracted from I_p, to obtain a high-frequency residual containing edges and fine textures as well as noise:

$$F = I_p - \mathrm{Up}(F_{low}) \quad (4)$$

Here F_{low} denotes the low-frequency image component. A subsequent 3×3convolution with Swish activation converts F_{high} into a 64-channel feature map F_{high}.

Frequency Fusion. The two branches are concatenated and passed through a 1×1 convolution to mix channels and reduce redundancy:

$$F_0 = \psi(\text{concat}(F_{low}, F_{high})) \quad (5)$$

Here $\boldsymbol{\psi}(\cdot)$outputs 64 channels. This fused tensor provides the base representation for higher-level feature extraction.

3.4 | Dual Branch Feature Extraction

To jointly model local patterns and long-range context, DIAA adopts a dual-branch design.

Multiscale CNN Branch. The multiscale branch applies two parallel dilated convolutions on F_0:

- 3×3 with a dilation rate of 1,
- 3×3 with a dilation rate of 2.

The resulting 32-channel feature maps are concatenated and processed by a 1 × convolution to yield 64 multiscale features. This branch effectively enlarges the receptive field while maintaining resolution, enabling robust modeling of structures across different spatial scales.

Transformer-CNN Branch. In parallel, the transformer-CNN branch first applies a standard 3×3 convolution with Swish activation to generate 64 base features. A dedicated spatial attention module computes a spatial attention map by concatenating per-pixel average and max responses across channels, processing them through a 7×7 convolution and sigmoid activation. The attention map is multiplied by the base features, yielding attention-weighted features that emphasize informative regions, such as edges and illuminated structures.

The base and attention-weighted features are concatenated to form a 128-channel tensor.

Branch Fusion. The outputs from the multiscale CNN (F_{ms}) and transformer-CNN (F_{tr}) Branches are finally concatenated:

$$F_{mb} = \text{concat}(F_{ms}, F_{tr}) \quad (6)$$

yielding a 192-channel multi-branch feature $(\boldsymbol{F_{mb}})$ representation that feeds the noise-aware residual stack.

3.5 | Noise Aware Residual Refinement

The model applies three identical noise-aware residual blocks to $F_{mb}$for progressive denoising and enhancement. Each block receives an input tensor $F \in \mathbb{R}^{H \times W \times C}$and processes it as follows:

Main Path. The main path consists of two 3×3 convolutions, each followed by batch normalization and Swish activation. This sequence extracts informative features while preserving stability and gradient flow.

Noise Path. A parallel noise path approximates the noise component by applying two consecutive 3×3 convolutions (without normalization) to the same input. The output is a noise estimate N with C channels.

Feature–Noise Separation. The block subtracts the estimated noise from the main path features:

$$F_{clean} = F_{main} - N \quad (7)$$

This operation explicitly encourages the network to disentangle noise from structural content.

Channel and Spatial Attention. To further refine features, a squeeze-and-excitation (SE) module produces channel weights by global average pooling F_{clean}, passing the pooled vector through two fully connected layers with ReLU and sigmoid activations, and scaling each channel accordingly. A spatial attention module then aggregates information across channels via average and max pooling, applies a 7×7convolution and sigmoid, and multiplies the result with the SE-weighted tensor. This combination adaptively emphasizes informative channels and spatial locations.

Residual Connection. Finally, a residual shortcut adds the refined features ($F_{refined}$) resulted from Channel and Spatial Attention blocks back to the block's input:

$$F_{out} = F + F_{refined} \tag{8}$$

Stacking three such blocks produce a robust representation that is both denoised and contrast-enhanced while preserving fine structures.

3.6 | Global–Local Frequency Fusion

After residual refinement, DIAA merges global context with local detail.

Global Context Path. A global average pooling layer aggregates features across spatial dimensions, producing a $1 \times 1 \times C$ descriptor capturing entire-image statistics. A 1×1 convolution with Swish activation transforms this descriptor, which is then resized via bilinear interpolation back to $H \times W$. This broadcast map represents low-frequency, scene-level information.

Local Detail Path. The local branch applies an additional 3×3 convolution with Swish activation directly on the residual-refined features. This branch focuses on contrast, edges, and textures.

Fusion and Channel Attention. The global and local tensors are concatenated and passed through a Frequency Aware block, implemented as a 1×1 convolution that mixes channels and reduces dimensionality to 64. A channel attention module, such as SE, recalibrates channel responses. The resulting tensor F_{fuse} simultaneously encodes global illumination behavior and local structural detail.

3.7 | Residual Enhancement and Reconstruction

The reconstruction head uses residual learning to generate the final enhancement.

First, a 3×3 convolution with tanh activation transforms F_{fuse} into a 3-channel feature map. A second 3×3 convolution with tanh activation predicts a residual map $R \in \mathbb{R}^{H \times W \times 3}$ that models both brightening and darkening corrections. A global residual addition combines this map with an intermediate feature representation (or the preprocessed image, depending on implementation):

$$I_{enh} = F_{base} + R \tag{9}$$

A final sigmoid activation constrains I_{enh} to $[0,1]$, yielding the enhanced RGB output I_{out}.

3.8 | Metric Guided Hybrid Loss

We utilized a hybrid loss function that integrates several complementary terms.

1. Pixel Fidelity (MSE and L1) focuses on two metrics: Mean Squared Error (MSE), which ensures global accuracy, and L1 loss [15], which enhances edge sharpness and robustness to outliers. L1 loss, or Mean Absolute Error (MAE), is calculated between actual and predicted images, promoting sparsity and minimizing absolute differences. This approach ensures the denoised photos closely match the original images while preserving edges and resisting outlier influence. MSE loss and L1 loss can be expressed as follows:

   Mean squared error $L_{\text{MSE}} = \frac{1}{n}\sum_{i=0}^{n}\left(S_{output}^{(i)} - I_{clean}^{(i)}\right)^2$ (10)

   L1 loss $L_{\text{MAE}} = \frac{1}{n}\sum_{i=0}^{n}|S_{output}^{(i)} - I_{clean}^{(i)}|$ (11)

   Where $S_{output}^{(i)}$, $I_{clean}^{(i)}$, represent the output image and corresponding target images.

2. **Structural Similarity (SSIM)** [16] promotes the maintenance of luminance, contrast, and structural patterns.

$$L_{\text{SSIM}} = 1 - SSIM(S_{output}^{(i)} - I_{clean}^{(i)}) \quad (12)$$

3. Perceptual Loss extracts High-level features from both output and ground truth using a pretrained VGG 19 network (without the classification head). The mean squared difference between matching feature maps is used to calculate the perceptual loss, which penalizes textures and structures that seem improbable.

4. Illumination Smoothness Loss calculates the Spatial gradients of I_{out} in both horizontal and vertical directions. The average absolute gradient magnitude promotes smooth light transitions while permitting edges in reflectance.

5. Color Constancy Loss calculates Channel means μ_R, μ_G, μ_B over I_{out}. This term suppresses undesirable color casts and maintains white balance consistency by penalizing deviations from these means.

6. Exposure Control Loss aids in preventing underexposure or overexposure by setting a target exposure level, thereby constraining the global mean intensity of the output. This approach helps the network achieve visually appealing brightness levels.

7. Knowledge Distillation loss [17] involves utilizing a high-capacity teacher model to enhance supervision through distillation loss. This process generates prediction tensors from both the teacher and student models, scales them by a temperature parameter, and applies the softmax function. The distillation loss can be expressed as:

$$\text{Distillation Loss} = \text{KL}\big(\text{Softmax}(p_t/T) \parallel \text{Softmax}(p_s/T)\big) \quad (13)$$

Where p_t , p_s represent the teacher and student predictions, and T is a temperature variable.

The overall loss combines the above terms with an optional distillation:

$$L_{\text{total}} = L_{\text{SSIM}} + L_{\text{perc}} + L_{\text{illum}} + L_{\text{color}} + L_{\text{exp}} + L_{\text{KD}} \quad (14)$$

In practice, this composite custom loss steers the network toward outputs that are numerically accurate, structurally faithful, perceptually convincing, and visually natural in terms of color and exposure, while leveraging teacher supervision when provided.

4 | Experiments

4.1 | Datasets and Data Preparation

The performance of the DIAA model is assessed using the LOL dataset [37] combined the low-light and normal light images that constitute the LOLv1 challenge, which is derived from the public distribution on the Kaggle platform. In the DIAA model, the training data consists of 485 pairs of low-light and corresponding normal-light images, and the test data consists of 15 pairs. All pairs are aligned . In contrast to traditional networks, which process images at the original size, the pairs are split into multiple training examples. Each picture is scanned at 100%, 90%, 80%, and 70% of the original image size. For each scale, the images are segmented into 64x64 non-overlapping patches using a window of the same size. For each low-light patch, there exists an equivalent patch from an image taken in normal light.

To reduce overfitting and improve robustness to viewpoint transformations, some fundamental geometric transformations are applied during learning. Every patch for the respective images undergoes flips and rotations at 0°, 90°, 180°, and 270°, yielding eight variants for content and lighting. These are further converted into matrices, and 8-bit signed values are changed to 32-bit float values. Before beginning the training process, the total number of patches is ensured to be divisible by the batch size. During training, the data generator produces an infinite number of low-light image batches as inputs and normal-light image targets. This patch-based, multi-scale, and extensively enhanced method significantly increases the quantity and diversity of the LOL dataset, thereby improving the network's generalization power.

4.2 | Implementation details

The proposed DIAA model is implemented in Keras/TensorFlow and trained on the LOL paired dataset using the Kaggle GPU environment with 16 GB of GPU memory. Before training, any previous session is cleared, and checkpoints are scanned so that training can resume from the last saved epoch, if available. The network is optimized using the Adam optimizer with an initial learning rate of 1×10^{-4}, $\beta_1 = 0.7$, and $\beta_2 = 0.999$, and is trained with a custom hybrid loss that includes a distillation term (distillation weight 0.3, temperature 0.8).

Training is performed using a patch-based data generator that produces normalized 64×64 low-/normal-light pairs with on-the-fly augmentation, for 2000 steps per epoch for 10 epochs. A set of callbacks is employed during training, including model checkpointing at every epoch, CSV logging of training statistics, and a ReduceLROnPlateau scheduler that monitors PSNR and reduces the learning rate when performance saturates, ensuring stable convergence on the Kaggle platform.

5 | Results and Analysis

All experiments in this work are conducted on this paired LOL dataset. Quantitative evaluation is performed on the held-out LOL test pairs using standard full-reference image quality measures. Peak Signal to Noise Ratio and the Structural Similarity Index were used for similarity to the normal images, and the Mean Absolute Error for absolute accuracy. The proposed approach is compared to several baselines and advanced techniques, along with relevant tables and figures. The results are also analyzed to present the merits and demerits of the new method, the findings from the experimental part, and the variables affecting the performance of the proposed models.

5.1 |Quantitative Results

The performance of the models is evaluated by using standard quantitative evaluation metrics, such as PSNR, SSIM, and MAE. These evaluation metrics are then compared to the performance metrics obtained by standard models and cutting-edge models on the same dataset. The performance comparison suggests that our model outperforms standard models and achieves performance comparable to cutting-edge models. Table 1 below highlights the improvements achieved by the models on PSNR, SSIM, and MAE.

Table 1. Quantitative Comparisons on the LOL dataset.

Method	PSNR (dB) ↑	SSIM ↑	MAE ↓
LIME [27]	16.899	0.472	0.137
Zero-DCE [31]	14.863	0.569	0.218
SCI [32]	14.457	0.565	0.193
URetinex-Net [28]	21.321	0.833	0.093
EnlightenGAN[34]	18.082	0.666	0.135
KinD [29]	16.834	0.772	0.171
Proposed	21.937	0.839	0.065

Our experiments show that the DIAA model significantly outperforms both traditional and modern low-light enhancement techniques on the LOL dataset. It achieves the highest PSNR score of 21.94 dB and SSIM of 0.839, while maintaining the lowest error rate (MAE of 0.065). What this means in practical terms is that DIAA produces enhanced images that look much closer to the original, well-lit reference photos in both clarity and structural integrity. What's particularly impressive is how DIAA surpasses even the previous top performer, URetinex-Net (which scored 21.32 dB PSNR, 0.833 SSIM, and 0.093 MAE). This improvement

demonstrates the real advantage of integrating illumination modeling, frequency-aware processing, and knowledge distillation into a single cohesive system.

Older methods like LIME struggle with lower-quality scores and higher error rates, revealing their inability to effectively handle noise, extreme darkness, and color issues simultaneously. While lightweight models such as Zero-DCE and SCI run faster, they fall short of DIAA's performance across the board—their predefined rules don't adapt well to the real-world lighting challenges present in the LOL dataset. Newer approaches, such as the GAN-based EnlightenGAN and the Retinex-inspired KinD, produce more visually appealing results than classical methods. However, their measured performance still can't match DIAA's, particularly in preserving the original scene's structure and details.

In summary, the quantitative improvements in PSNR, SSIM, and MAE substantiate that the proposed model performs superiorly in global brightness correction and meticulous detail restoration compared to current methodologies.

5.2 | Qualitative Results

Beyond just looking at numbers, we wanted to see how the enhanced images look to the human eye. We carefully examined sample images from the LOL dataset, comparing our model's results side-by-side with those from both established and cutting-edge enhancement techniques. We go beyond visual inspection and check RGB color histograms for a fuller assessment of quality. Those histograms can reveal whether each method truly recovers natural brightness and color balance, or if something is off.

As shown in Figure 3, our approach produces images that are noticeably more explicit, with better contrast and significantly less graininess or noise. Even more telling, the color histograms align much more closely with those of the original well-lit reference photos. In short, our model isn't just brightening images-it's restoring colors and lighting in a way that looks natural.

The visual comparisons in Figure 3 highlight apparent differences between existing methods and the proposed DIAA model. Traditional approaches such as LIME and lightweight learning-based methods like Zero-DCE and SCI often fail to strike a balance between brightness and natural appearance, frequently leaving regions underexposed, washing out details, or amplifying noise. Methods including EnlightenGAN, URetinex-Net, and KinD generally recover more structure and produce globally brighter outputs. Still, they tend to introduce color shifts, warm casts, halos, or exaggerated edges around high-contrast boundaries.

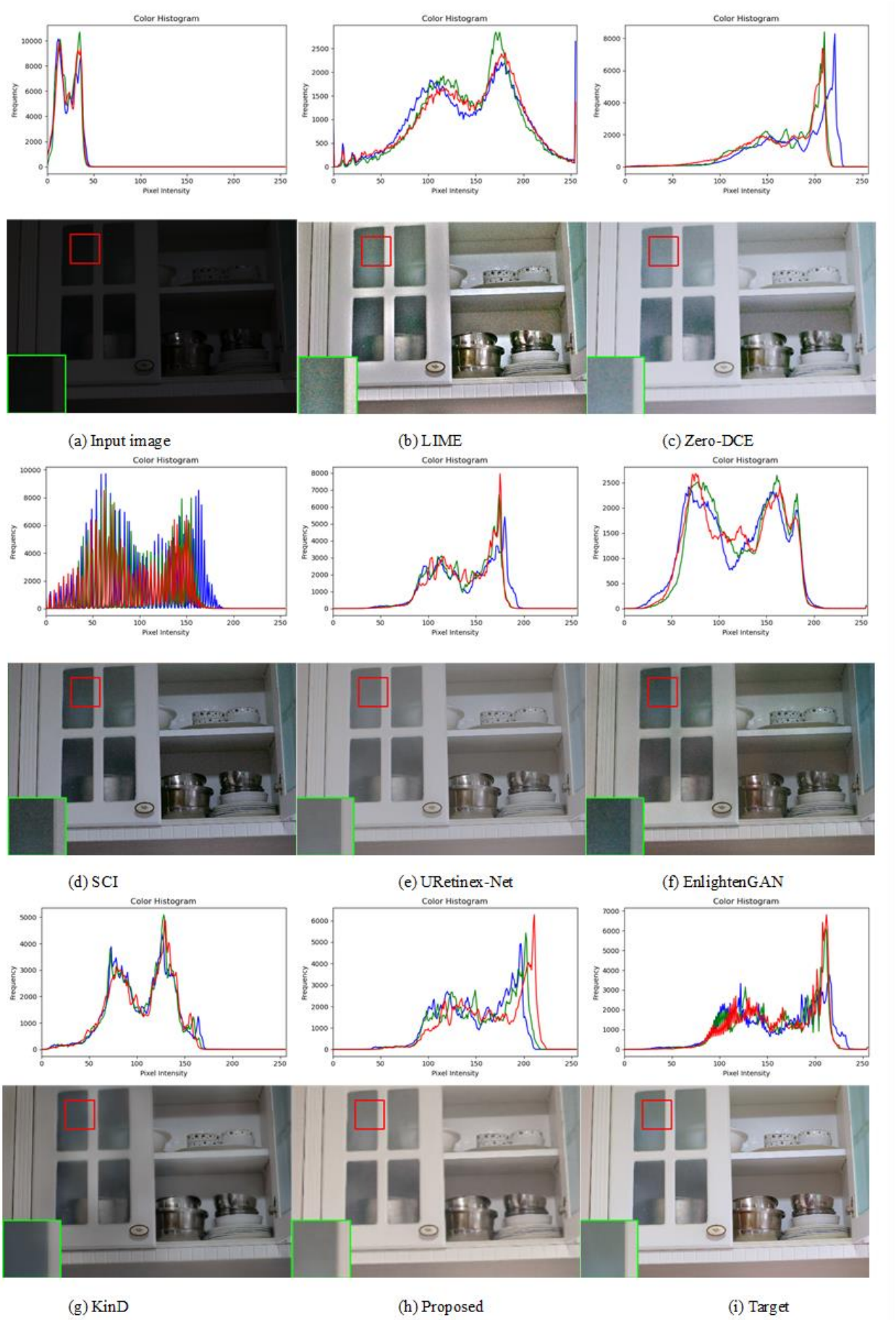

Figure 3. Visual comparison with color histogram using state-of-the-art algorithms on the LOLv1 dataset.

In contrast, the proposed DIAA model effectively enhances dark regions while maintaining color integrity and preventing overexposure. It improves local contrast, uncovers details in shadows, and preserves textures without excessive smoothing or noise amplification. Under various low-light conditions, DIAA outputs closely match those from normal-light images, indicating that the employed noise-aware residual blocks and frequency fusion successfully separate structure from noise, resulting in superior visual quality. These results affirm that the proposed architecture, which uses a hybrid loss and ResNet50-based distillation, provides a reliable and efficient solution for low-light image enhancement on realistic datasets such as LOL.

6 | Conclusion and Future Work

The proposed DIAA framework demonstrates strong capability in enhancing low-light images while suppressing noise and preserving perceptual quality. Across quantitative evaluations, including PSNR, SSIM, and MAE on the LOL dataset, the model consistently outperforms several low-light enhancement methods, indicating better reconstruction fidelity and structural preservation. Visual inspection further confirms that DIAA produces images with more natural illumination, greater detail precision, and fewer artifacts than competing approaches .

These results suggest that DIAA is a promising solution for low-light image enhancement, particularly in scenarios where both visual quality and robustness to noise are critical. Future work will focus on further reducing the computational complexity of the student network to facilitate deployment on more constrained hardware, as well as exploring alternative loss formulations and transfer learning strategies to improve generalization across different low-light domains and imaging conditions.

Author Contribution

All authors contributed equally to this work.

Funding

This research has no funding source.

Data Availability

The dataset used in this study is publicly available in the LOL dataset repository at https://daooshee.github.io/BMVC2018website/.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

References

[1] Hou, Y., & Yang, B. (2024). Semantic attention guided low-light image enhancement with multi-scale perception. Journal of visual communication and image representation, 103, 104242. DOI:https://doi.org/10.1016/j.jvcir.2024.104242

[2] Cai, J., Gu, S., & Zhang, L. (2018). Learning a deep single image contrast enhancer from multi-exposure images. IEEE transactions on image processing. DOI:10.1109/TIP.2018.2794218

[3] Celik, T., & Tjahjadi, T. (2011). Contextual and variational contrast enhancement. IEEE transactions on image processing. DOI:10.1109/TIP.2011.2157513

[4] Thomas, G., Flores-Tapia, D., & Pistorius, S. (2011). Histogram specification: a fast and flexible method to process digital images [presentation]. IEEE transactions on instrumentation and measurement. DOI: 10.1109/TIM.2010.2089110

[5] Lee, C., Kim, C. S., & Lee, C. (2013). Contrast enhancement based on layered difference representation of 2D histograms. IEEE transactions on image processing. DOI:10.1109/TIP.2013.2284059

[6] Arici, T., Dikbas, S., & Altunbasak, A. (2009). A histogram modification framework and its application for image contrast enhancement. IEEE transactions on image processing. DOI:10.1109/TIP.2009.2021548

[7] Huang, S. C., Cheng, F. C., & Chiu, Y. S. (2013). Efficient contrast enhancement using adaptive gamma correction with weighting distribution. IEEE transactions on image processing. DOI:10.1109/TIP.2012.2226047
[8] Li, X., Liu, M., & Ling, Q. (2024). Pixel-Wise Gamma Correction Mapping for Low-Light Image Enhancement. IEEE transactions on circuits and systems for video technology. DOI:10.1109/TCSVT.2023.3286802
[9] Ozturk, N., & Ozturk, S. (2024). Efficient and natural image fusion method for low-light images based on active contour model and adaptive gamma correction. Multimedia tools and applications. DOI:10.1007/s11042-023-17141-8
[10] Rao, K., Bansal, M., & Kaur, G. (2022). Retinex-Centered Contrast Enhancement Method for Histopathology Images with Weighted CLAHE. Arabian journal for science and engineering. DOI:10.1007/s13369-021-06421-w
[11] Wang, P., Wang, Z., Lv, D., Zhang, C., & Wang, Y. (2021). Low illumination color image enhancement based on Gabor filtering and Retinex theory. Multimedia tools and applications. DOI:10.1007/s11042-021-10607-7
[12] Rafael C. Gonzalez, R. E. W. (n.d.). Digital Image Processing.
[13] Li, C., Guo, J., Porikli, F., & Pang, Y. (2018). LightenNet: A Convolutional Neural Network for weakly illuminated image enhancement. Pattern recognition letters. DOI:10.1016/j.patrec.2018.01.010
[14] Kim, Y. T. (1997). Contrast enhancement using brightness preserving bi-histogram equalization. IEEE transactions on consumer electronics. DOI:10.1109/30.580378
[15] Chen, S. Der, & Ramli, A. R. (2004). Preserving brightness in histogram equalization based contrast enhancement techniques. Digital signal processing: a review journal. DOI:10.1016/j.dsp.2004.04.001
[16] Ooi, C. H., & Isa, N. A. M. (2010). Quadrants dynamic histogram equalization for contrast enhancement. IEEE transactions on consumer electronics. DOI:10.1109/TCE.2010.5681140
[17] Kim, M., & Chung, M. G. (2008). Recursively separated and weighted histogram equalization for brightness preservation and contrast enhancement. IEEE transactions on consumer electronics. DOI:10.1109/TCE.2008.4637632
[18] Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., … Zuiderveld, K. (1987). ADAPTIVE HISTOGRAM EQUALIZATION AND ITS VARIATIONS. Computer vision, graphics, and image processing. DOI:10.1016/S0734-189X(87)80186-X
[19] Shi, Z., Feng, Y., Zhao, M., Zhang, E., & He, L. (2020). Normalised gamma transformation-based contrast-limited adaptive histogram equalisation with colour correction for sand-dust image enhancement. IET image processing. DOI:10.1049/iet-ipr.2019.0992
[20] Wadud, M. A. Al, Kabir, M. H., Dewan, M. A. A., & Chae, O. (2007). A dynamic histogram equalization for image contrast enhancement. IEEE transactions on consumer electronics. DOI:10.1109/TCE.2007.381734
[21] Ibrahim, H., & Kong, N. S. P. (2007). Brightness preserving dynamic histogram equalization for image contrast enhancement. IEEE transactions on consumer electronics. DOI:10.1109/TCE.2007.4429280
[22] Cao, X., & Yu, J. (2024). LLE-NET: A Low-Light Image Enhancement Algorithm Based on Curve Estimation. Mathematics.
[23] Jobson, D. J., Rahman, Z. U., & Woodell, G. A. (1997). Properties and performance of a center/surround retinex. IEEE transactions on image processing. DOI:10.1109/83.557356
[24] Jobson, D. J., Rahman, Z. U., & Woodell, G. A. (1997). A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE transactions on image processing. DOI:10.1109/83.597272
[25] Wang, S., Zheng, J., Hu, H. M., & Li, B. (2013). Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE transactions on image processing. DOI:10.1109/TIP.2013.2261309
[26] Fu, X., Zeng, D., Huang, Y., Zhang, X. P., & Ding, X. (2016). A weighted variational model for simultaneous reflectance and illumination estimation [presentation]. Proceedings of the ieee computer society conference on computer vision and pattern recognition. DOI: 10.1109/CVPR.2016.304
[27] Guo, X., Li, Y., & Ling, H. (2017). LIME: Low-light image enhancement via illumination map estimation. IEEE transactions on image processing. DOI:10.1109/TIP.2016.2639450
[28] Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., & Jiang, J. (2022). URetinex-net: retinex-based deep unfolding network for low-light image enhancement [presentation]. Proceedings of the ieee computer society conference on computer vision and pattern recognition (Vol. 2022-June, pp. 5891–5900). DOI: 10.1109/CVPR52688.2022.00581
[29] Zhang, Y., Zhang, J., & Guo, X. (2019). Kindling the darkness: a practical low-light image enhancer [presentation]. MM 2019 - proceedings of the 27th acm international conference on multimedia. DOI: 10.1145/3343031.3350926
[30] Zhang, Y., Guo, X., Ma, J., Liu, W., & Zhang, J. (2021). Beyond Brightening Low-light Images. International journal of computer vision. DOI:10.1007/s11263-020-01407-x
[31] Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement [presentation]. Proceedings of the ieee computer society conference on computer vision and pattern recognition. DOI: 10.1109/CVPR42600.2020.00185
[32] Ma, L., Ma, T., Liu, R., Fan, X., & Luo, Z. (2022). Toward fast, flexible, and robust low-light image enhancement [presentation]. Proceedings of the ieee computer society conference on computer vision and pattern recognition. DOI: 10.1109/CVPR52688.2022.00555
[33] Liu, R., Ma, L., Zhang, J., Fan, X., & Luo, Z. (2021). Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement [presentation]. Proceedings of the ieee computer society conference on computer vision and pattern recognition. DOI: 10.1109/CVPR46437.2021.01042
[34] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., … Wang, Z. (2021). EnlightenGAN: Deep Light Enhancement without Paired Supervision. IEEE transactions on image processing. DOI:10.1109/TIP.2021.3051462

[35] Wang, Y., Wan, R., Yang, W., Li, H., Chau, L. P., & Kot, A. (2022). Low-light image enhancement with normalizing flow [presentation]. Proceedings of the 36th aaai conference on artificial intelligence, aaai 2022. DOI: 10.1609/aaai.v36i3.20162

[36] Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., & Zhang, Y. (2023). Retinexformer: one-stage retinex-based transformer for low-light image enhancement [presentation]. Proceedings of the ieee international conference on computer vision. DOI: 10.1109/ICCV51070.2023.01149

[37] Wei, C., Wang, W., Yang, W., & Liu, J. (2019). Deep retinex decomposition for low-light enhancement [presentation]. British machine vision conference 2018, bmvc 2018.