

Article

DBENet: Dual-Branch Brightness Enhancement Fusion Network for Low-Light Image Enhancement

Yongqiang Chen ¹, Chenglin Wen ^{2,*}, Weifeng Liu ¹  and Wei He ¹

¹ School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China; 210611029@sust.edu.cn (Y.C.); liuwf@sust.edu.cn (W.L.); 220612047@sust.edu.cn (W.H.)

² School of Automation, Guangdong University of Petrochemical Technology, Maoming 525000, China

* Correspondence: wenc1@gdupt.edu.cn

Abstract: In this paper, we propose an end-to-end low-light image enhancement network based on the YCbCr color space to address the issues encountered by existing algorithms when dealing with brightness distortion and noise in the RGB color space. Traditional methods typically enhance the image first and then denoise, but this amplifies the noise hidden in the dark regions, leading to suboptimal enhancement results. To overcome these problems, we utilize the characteristics of the YCbCr color space to convert the low-light image from RGB to YCbCr and design a dual-branch enhancement network. The network consists of a CNN branch and a U-net branch, which are used to enhance the contrast of luminance and chrominance information, respectively. Additionally, a fusion module is introduced for feature extraction and information measurement. It automatically estimates the importance of corresponding feature maps and employs adaptive information preservation to enhance contrast and eliminate noise. Finally, through testing on multiple publicly available low-light image datasets and comparing with classical algorithms, the experimental results demonstrate that the proposed method generates enhanced images with richer details, more realistic colors, and less noise.

Keywords: low-light image enhancement; YCbCr space; dual-branch network; feature fusion



Citation: Chen, Y.; Wen, C.; Liu, W.; He, W. DBENet: Dual-Branch Brightness Enhancement Fusion Network for Low-Light Image Enhancement. *Electronics* **2023**, *12*, 3907. <https://doi.org/10.3390/electronics12183907>

Academic Editor: Eva Cernadas

Received: 17 August 2023

Revised: 11 September 2023

Accepted: 12 September 2023

Published: 16 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, with the continuous improvement of computer hardware and algorithms, artificial intelligence has made remarkable progress in various fields, such as image recognition [1], object detection [2], semantic segmentation [3], and autonomous driving [4]. However, these technologies are mainly based on the assumption that images are captured under good lighting conditions, and there are few discussions on target recognition and detection technologies under weak illumination conditions such as insufficient exposure at night, unbalanced exposure, and insufficient illumination. Due to the low brightness, poor contrast, and color distortion of images and videos captured at night (example shown in Figure 1), the effectiveness of visual systems, such as object detection and recognition, is seriously weakened. Enhancing the quality of images captured under low-light conditions via low-light image enhancement (LLIE) can help improve the accuracy and effectiveness of many imaging-based systems. Therefore, LLIE is an essential technique in computer vision applications.

Currently, various methods have been proposed for LLIE, including histogram equalization (HE) [5,6], non-local means filtering [7], Retinex-based methods [8,9], multi-exposure fusion [10–12], and deep-learning-based methods [13–15], among others. While these approaches have achieved remarkable progress, two main challenges impede their practical deployment in real-world scenarios. First, it is difficult to handle extremely low illumination conditions. Deep-learning-based methods show satisfactory performance in slightly low-light images, but they perform poorly in extremely dark images. Additionally, due to the low signal-to-noise ratio, low-light images are usually affected by strong noise. Noise

pollution and color distortion also bring difficulties to this task. Most of the previous studies on LLIE have focused on dealing with one of the above problems.



Figure 1. The comparison effect of various images taken in different scenes. From left to right, these images are derived from LOL, VE-LOL, LIME, and MEF datasets, respectively.

To explore the above problems, we counted the differences between 500 pairs of real low-/normal-light image pairs captured in the VE-LOL dataset in different color spaces and channels, as shown in Figure 2. In the RGB color space, all three channels exhibit significant degradation. However, in the YCbCr color space, the chrominance channels show higher PSNR and SSIM values compared to the luminance channel, indicating more severe image quality loss in the luminance channel. The inherent characteristics of the YCbCr color space indicate that the difference in luminance primarily resides in the Y channel, while the Cb and Cr channels are more susceptible to noise contamination. To achieve the goal of decoupling luminance distortion and noise interference, it is possible to employ channel-wise processing to handle different channels more appropriately. Therefore, in low-light image enhancement tasks, compared to the RGB color space, the YCbCr color space provides a favorable potential candidate space for separating luminance distortion and noise interference.

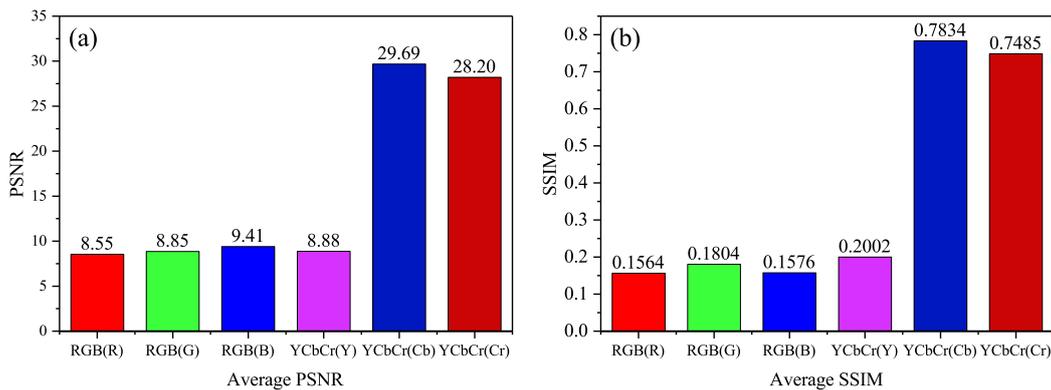


Figure 2. The difference between low-light images and normal images in RGB space and YCbCr space under the VE-LOL dataset. (a) The average PSNR values for each channel; (b) The average SSIM values for each channel.

In summary, the main contributions of this article are as follows:

- We propose a new hierarchical structure (DBENet) for enhancing low-light conditions in the real world. This framework includes networks for enhancing illumination maps, denoising chromatic information, and feature map fusion, respectively;
- We employed a CNN branch to predict the gamma matrix and utilized nonlinear mapping to regulate brightness variations, effectively suppressing overexposure during the enhancement process;

- Our method outperforms existing techniques on benchmark datasets, achieving significant improvements in evaluation metrics such as MAE, PSNR, SSIM, LPIPS (reference), and NIQE (no-reference), demonstrating its superior efficiency.

The rest of this paper is as follows: Section 2 introduces the proposed network framework. Section 3 explains the loss function used in each component. Section 4 presents the evaluation of our method via subjective and objective assessments of multiple datasets. Sections 5 and 6 are dedicated to the discussion and conclusion, respectively

2. Related Works

In general, image enhancement methods can be roughly divided into two categories: non-learning-based methods and learning-based methods.

2.1. Non-Learning-Based Methods

LLIE plays an irreplaceable role in recovering the intrinsic colors and details, as well as compressing noise in low-light images. In the following, we provide a comprehensive review of previous work on low-light image enhancement. Traditional LLIE methods encompass techniques such as tone mapping [16], gamma correction [17], histogram equalization [18], and those based on the Retinex theory [19–22]. Tone mapping is used to create more detailed, colorful, and high-contrast images while maintaining a natural appearance. However, linear mapping can lead to the loss of information in bright and dark areas. Gamma correction employs nonlinear tone mapping to handle the shadows and highlights in image signals, but selecting global parameters can be difficult and may result in overexposure or underexposure. Histogram equalization enhances image contrast by transforming the histogram, but it may yield unsatisfactory results in certain local regions. Adaptive histogram equalization [23] can map the histogram of local regions to a simpler distribution for improved effects. The Retinex theory [24] is a computational theory that simulates human visual perception and can achieve color constancy, color enhancement, and high dynamic range compression. However, there is still room for improvement in its processing mechanisms and universality, and its effectiveness may vary in different scenarios. In general, traditional model-based methods heavily rely on manually designed priors or statistical models, which may limit their applications.

2.2. Learning-Based Methods

In the field of LLIE, methods based on deep learning have currently become the mainstream research direction. LLNet [25] represents a seminal contribution from the LLIE group, which focuses on contrast enhancement and denoising via a depth autoencoder-based approach. However, it is worth noting that this work does not explore the intricate relationship between real-world illumination and noise, consequently leading to persistent issues such as residual noise and excessive smoothing. In contrast, Chen et al. [26] introduced Retinex-Net, a method that decomposes the input image into a reflectance map and an illumination map. It enhances the illumination map using a deep neural network for low-light conditions and then applies BM3D [27] for denoising, while Retinex-Net effectively enhances brightness and image details, it tends to suffer from inadequate image smoothing and severe color distortion. Lv et al. [28] proposed a comprehensive end-to-end multi-branch enhancement network (MBLEN) encompassing feature extraction, enhancement, and fusion modules to boost the performance of LLIE. Drawing inspiration from super-resolution reconstruction techniques, UTVNet [29] and URetinex [30] introduced an adaptive unfolding network tailored for robustly denoising and enhancing low-light images. Another notable approach by Wang et al. [31] introduces a two-stage Fourier-based LLIE network, FourLLIE. This method enhances the brightness of low-light images by estimating amplitude transformation in the Fourier space. Furthermore, it leverages a signal-to-noise ratio (SNR) map to provide a priori information regarding global Fourier frequencies and local spatial details for image restoration. Notably, FourLLIE is both lightweight and highly effective in terms of enhancement.

Recently, zero-shot-learning-based methods has garnered substantial attention due to their efficiency, cost-effectiveness, and ability to leverage limited image data. For instance, Liu et al. [32] introduced Retinex-based Unrolling with Architecture Search (RUAS) and devised a collaborative reference-free learning strategy to discover low-light prior architectures from a compact search space. Guo et al. [33] presented Zero-DCE, a technique employing an intuitive nonlinear curve mapping. Subsequently, they improved upon this method with Zero-DCE++ [34], which is faster and lighter. However, it is important to note that Zero-DCE relies on multiple exposure training data and does not effectively address noise, especially in extreme enhancement scenarios. Zhu et al. [35] introduced RRDNet, a three-branch convolutional neural network designed for restoring underexposed images. RRDNet employs an iterative approach to decompose input images into their constituent parts: illumination, reflectance, and noise. This is achieved via the minimization of a customized loss function and the adjustment of the illumination map via gamma correction. The reconstructed reflectance and adjusted illumination map are then multiplied element-wise to generate the enhanced output. In another development, Ma et al. [36] proposed a learning framework called self-calibrating illumination (SCI) for rapid and adaptable enhancement in real-world low-illumination scene images. This method estimates a convergent illuminance map via a neural network and, following Retinex theory, divides the input low-illumination image element-wise with the estimated illuminance map to derive an enhanced reflectance map. It is worth noting that while SCI achieves a convergence of the illuminance map through iterations, it does not explicitly address noise interference in the process. PSENet [37] offers an unsupervised approach for extreme-light image enhancement, effectively addressing image enhancement challenges in both overexposure and underexposure scenarios.

3. The Proposed Network

In the third section, we first introduced our proposed DBENet and provided a more detailed explanation of the components we proposed in the following subsections.

The architecture of the proposed dual-branch enhancement network (DBENet) is shown in Figure 3. DBENet consists of two branches (CNN branch and U-Net branch) and a fusion module. The network follows a divide-and-conquer strategy, where the input image is transformed from the original RGB color space to the YCbCr color space for separate processing. The CNN branch handles the luminance component (Y) based on the nonlinear function. The encoder–decoder branch network processes the chrominance component ($CbCr$) starting from global features. Finally, the cascaded fusion features (Y_{res} and W_{res}) from both branches are fed into the fusion module to aggregate the enhanced image.

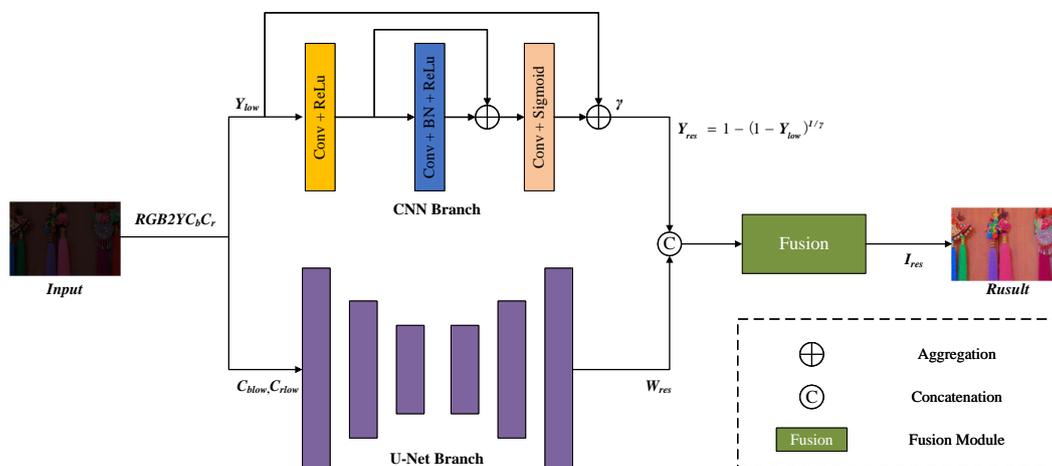


Figure 3. The proposed network structure framework diagram.

3.1. CNN Branch

The CNN branch based on the residual concept consists of three parts: the initial layer Conv + ReLU, the middle layer Conv + BatchNorm + ReLU, and the final layer Conv + Sigmoid. The convolutional kernel size is set uniformly to 3×3 with a dilation rate of 1, which enlarges the receptive field of the convolutional network and enhances the feature extraction ability without increasing the computational burden. The BatchNorm layer normalizes each channel to reduce inter-channel dependencies and accelerate network convergence. After obtaining the estimated gamma component γ through the network, we employ the gamma adjustment scheme [38] to enhance the visibility of details in both dark and bright regions. The nonlinear function is represented by the following equation:

$$Y_{res} = 1 - (1 - Y_{low})^{1/\gamma} \tag{1}$$

In Equation (1), Y_{res} represents the enhanced result, and γ and Y_{low} , respectively, denote the predicted gamma map and the separated luminance component of the original image. This function is designed to address the issue of overexposure that often occurs when enhancing results in the presence of non-uniform lighting and complex light sources in the original image. Unlike directly applying the gamma function to the original image, we draw inspiration from dehazing techniques and apply it to the inverted image to obtain the enhanced output. This approach arises from the shared characteristics of blurred and low-light images, which often exhibit low dynamic range and high noise levels. Therefore, dehazing techniques, such as using inverted images, can be employed to enhance and alleviate this concern.

Within the CNN branch, the process begins by normalizing the image to a 0–1 range. Subsequently, the network learns the intermediate parameter gamma for predicting the mapping function and, finally, computes the predicted result. As illustrated in Figure 4’s mapping curve, when the gamma value is less than 1, it brightens areas with underexposure, while gamma values greater than 1 darken areas with overexposure. The purpose of this function is to provide reasonable suppression, allowing the control and mitigation of the local intensity increase, while simultaneously enhancing the overall image quality.

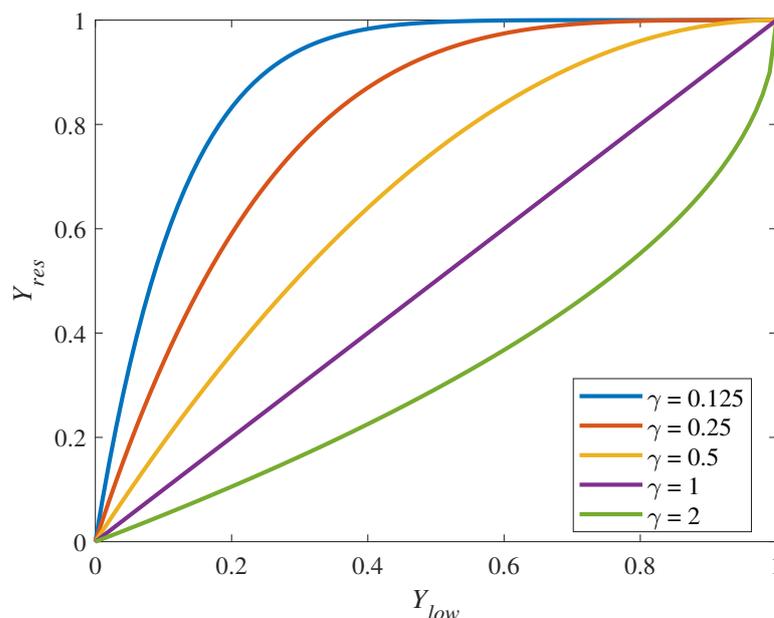


Figure 4. Function mapping curves corresponding to different γ values.

3.2. U-Net Branch

Due to the influence of the acquisition environment and equipment, low-illumination images often contain a lot of noise in dark areas. Noise will reduce image information and

image quality. In order to better dealing with low-light images, it is necessary to achieve better denoising and detail preservation effects.

In an effort to reveal the details while avoiding the increase in distortion, we propose a chromaticity denoising module. The module uses the chrominance channel of the low-illumination image to mainly reflect the chrominance information of the image, which can be represented as W . Since the color information distortion is often non-local, in order to obtain the global color information of the image, the classical U-Net network structure is used to enrich the spatial information by extracting features of different sizes so that the semantic information is more diverse. Through the encoder–decoder structure, the U-Net branch can capture context information at different scales. In addition, the introduction of skip connections enables U-Net [39] to make full use of feature information and restore details and boundaries, as shown in Figure 5. In the U-Net branch, the encoder expands the receptive field of convolution via layer-by-layer pooling operation. In the bottleneck layer of the network, the larger receptive field can extract the non-local chrominance information for contrast recovery, and the decoder expands the non-local information to the global via layer-by-layer upsampling.

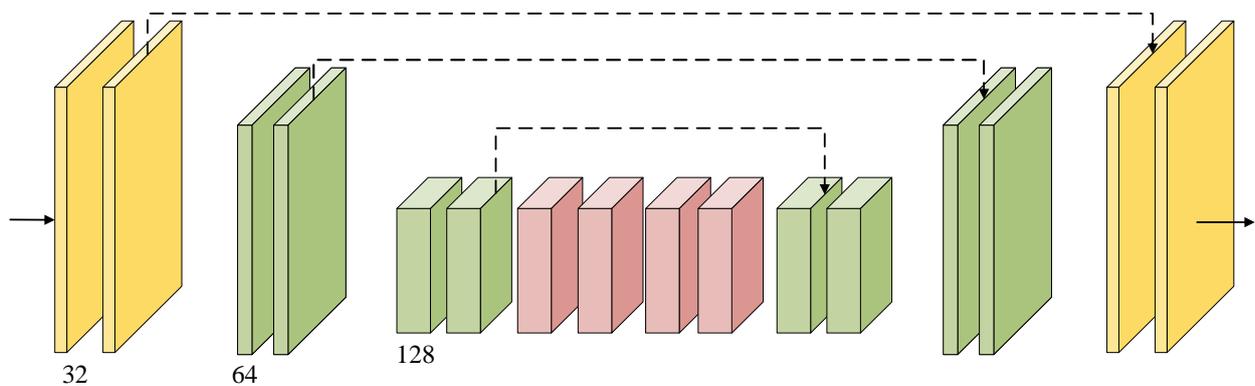


Figure 5. The structure of the U-Net Branch.

3.3. Fusion Module

In our method, we did not perform the corresponding transformation from YCbCr to RGB color space on the returned three components. Instead, we did not design a unique fusion rule but used a fusion module to generate the fused result I_{res} . As shown in Figure 6, the architecture of the fusion module consists of 10 layers, with Y_{res} and W_{res} concatenated as inputs. Each layer has a convolutional operation, followed by an activation function. The kernel size of all convolutional layers is set to 3×3 , with a stride of 1. The padding mode is set to “reflect” to prevent edge artifacts. No pooling layers are used to avoid information loss. The activation function in the first nine layers is LeakyReLU with a slope of 0.2, while the activation function in the last layer is Sigmoid. Furthermore, studies [40] have shown that building short connections between layers close to the input and layers close to the output can significantly deepen and effectively train neural networks. Therefore, in the first seven layers, dense connection blocks are utilized to improve information flow and performance. In these layers, shortcut direct connections are established in a feed-forward manner between each layer and all preceding layers, reducing the problem of vanishing gradients.

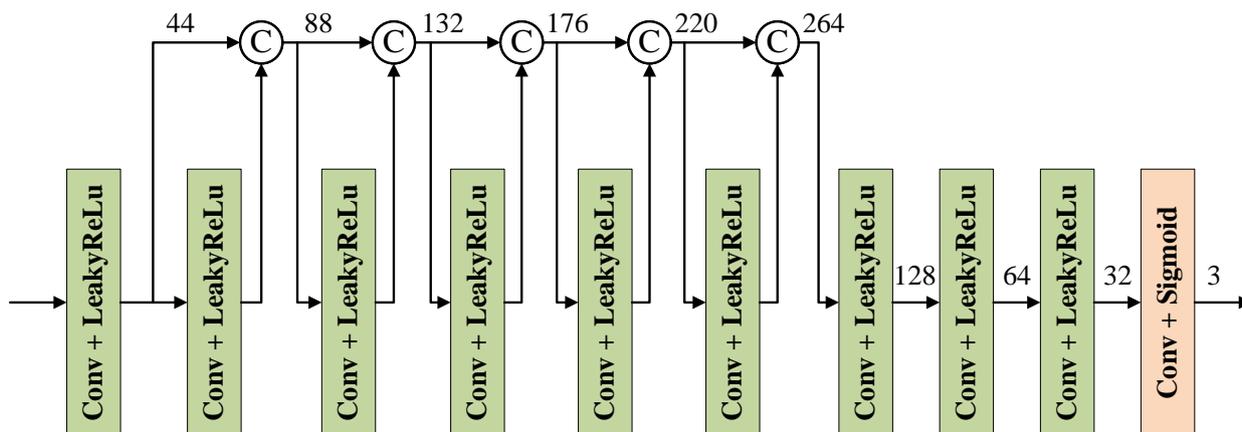


Figure 6. The structure of the fusion module. Numbers are the channels of corresponding feature maps.

3.4. Loss Function

During the training phase, due to the similar degradation patterns of the Cb and Cr chroma channels, for convenience, we use W to represent both the Cb and Cr channels simultaneously. The loss function of the entire network as follows:

$$\mathcal{L}_{Total} = \mathcal{L}_1(Y_{res}, Y_{high}) + \mathcal{L}_2(W_{res}, W_{high}) + \mathcal{L}_3(I_{res}, I_{high}) \tag{2}$$

Among these, I represents the output of the network, and Y and W represent the outputs of the CNN branch and the U-Net branch, respectively. The subscripts “res” and “high” indicate the enhanced result and the corresponding normal image.

In Equation (2), the three loss functions, \mathcal{L}_1 , \mathcal{L}_2 , and \mathcal{L}_3 , share the same form. Taking \mathcal{L}_1 as an example, we have $\mathcal{L}_1 = L_2 + L_{ssim}$. The two components represent the mean square error loss and the structural similarity loss function, respectively. The first term of the loss function aims to measure the reconstruction error, while the second term measures the differences in brightness, contrast, and structural similarity between the two images. Similarly, taking \mathcal{L}_1 as an example, the L_2 loss is defined as shown in Equation (3), while the definition of the L_{ssim} is presented in Equation (4).

$$L_2 = \left\| Y_{res} - Y_{high} \right\|_2^2 \tag{3}$$

$$L_{ssim} = 1 - SSIM(Y_{res} - Y_{high}) \tag{4}$$

where SSIM [41] is the structural similarity, the function is defined as follows:

$$SSIM(x, y) = \frac{(2 \times u_x \times u_y + c_1)(2 \times \sigma_{xy} + c_2)}{(u_x^2 + u_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{5}$$

4. Experimental Results and Analysis

In this part, we describe the experimental results and analysis in detail. First, we briefly introduce the experimental setting. Then, the qualitative and quantitative evaluation of paired and unpaired data sets is described. Finally, the experimental results are analyzed.

4.1. Experimental Settings

Parameter Settings: All experiments in this paper were conducted in the same configuration environment, i.e., training environment configuration: Ubuntu system, 32 GB RAM, and NVIDIA GeForce RTX3090 GPU. The network framework was constructed with the PyTorch framework and optimized using Admm [42] with parameters $\beta_1 = 0.9$, $\beta_2 = 0.99$, $\epsilon = 0.95$. In addition, the batch size was 16, the learning rate was 0.0002,

and the training sample size was uniformly adjusted to 256×256 . A total of 485 randomly selected paired images from the LOL dataset were used to train our model. The training epoch number was set to 3000.

Compared Methods: As for the low-light-level image intensifier, we conducted a visual evaluation of our proposed network on classic low-light image datasets (LOL and other datasets) and compared it with other state-of-the-art methods and available codes, including the traditional methods HE [5] and tone mapping [16], deep-learning-based methods Retinex-Net [26], RUAS [32], Zero-DCE [33], SCI [36], and RRDNet [35].

Evaluation Criteria: We employ quantitative image quality assessment metrics for comparative analysis to illustrate the effectiveness of the algorithms presented in this paper. To gauge the disparities in color, structural, and high-level feature similarity, we utilize MAE, PSNR, SSIM [41], LPIPS [43], and NIQE [44] as measurement indices. In addition, two paired data sets (LOL and VE-LOL) and two unpaired data sets (LIME and MEF) were selected for verification experiments to test their performance in image enhancement.

4.2. Subjective Visual Evaluation

Figures 7 and 8 show some representative results of the visual comparison of various algorithms. Figures 7 and 8 belong to the LOL and VE-LOL datasets, respectively. In Figure 7, it can be seen that HE has obvious image distortion and color distortion; Retinex-Net amplifies inherent noise, losing image details; SCI, Zero-DCE, and RRD-Net have weak brightness enhancement capabilities; tone mapping, RUAS, and our method perform extremely well in brightness and color aspects. From Figure 8, the enhanced results show that HE can significantly increase the brightness of low-light images. However, it applies contrast enhancement to each channel of RGB separately, causing color distortion. Retinex-Net significantly improves the visual quality of low-light images, but it overly smooths out details, enlarges noise, and even causes color deviation. Tone mapping can stretch the dynamic range of the image, but it still has insufficient enhancement for the grandstand seating section in the image. Although the image effect of RUAS is delicate and has no obvious noise interference, it does not successfully brighten the image in extremely dark areas (such as the central seat part). SCI and RRD-Net perform poorly in darker images and cannot effectively enhance low-light images. Zero-DCE can preserve the details of the image relatively completely, but the brightness enhancement is not obvious, and the color contrast of the image is significantly reduced. Compared with the ground truth, our method not only significantly improves brightness but also preserves colors and details to a large extent, thereby improving image quality.

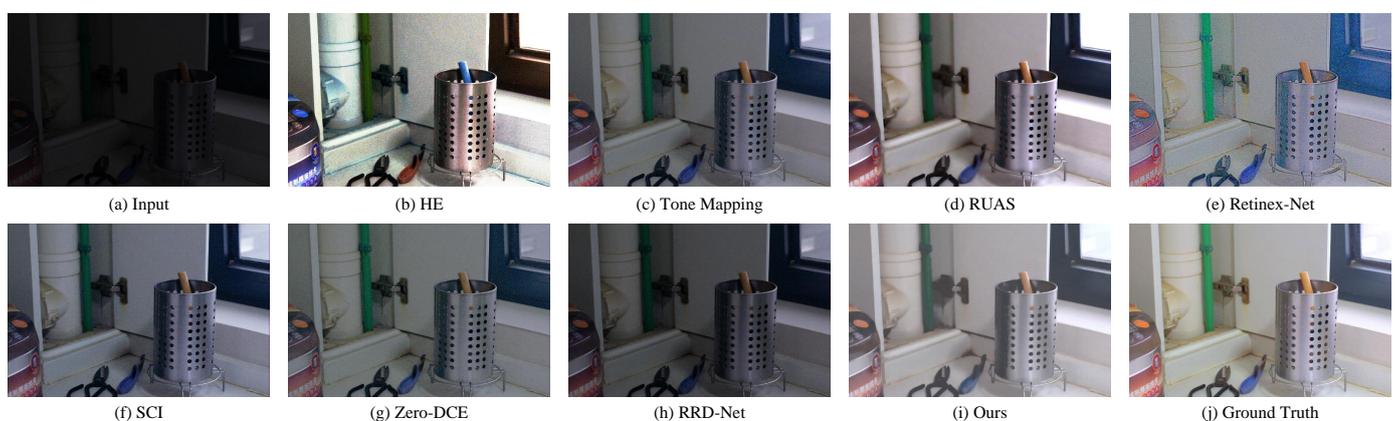


Figure 7. Visual comparisons of different approaches on the LOL benchmark.

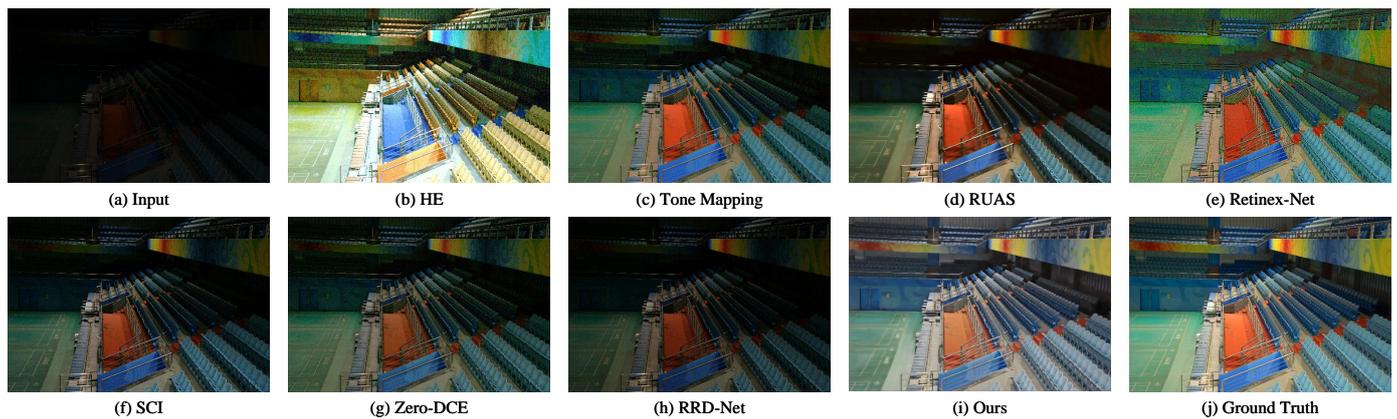


Figure 8. Visual comparisons of different approaches on the VE-LOL benchmark.

To comprehensively evaluate various algorithms, we also selected two unpaired benchmarks (LIME, MEF) for the verification experiments. As shown in Figures 9 and 10, we show the visual contrast effects produced via these cutting-edge methods under various benchmarks. From these enhancement results, it is evident that HE greatly improves the contrast of the image, but there is also a significant color shift phenomenon. Retinex-Net introduces visually unsatisfactory artifacts and noise. Tone mapping and RRD-Net can preserve image details, but the overall enhancement strength is not significant, and they fail to effectively enhance local dark areas. RUAS and SCI can effectively enhance low-contrast images, but during the enhancement process, they tend to excessively enhance originally bright areas, such as the sky and clouds in Figure 10, which are replaced by an overly enhanced white-ish tone. Among all the methods, Zero-DCE and our proposed method perform well on these two benchmarks, effectively enhancing image contrast while maintaining color balance and detail clarity.

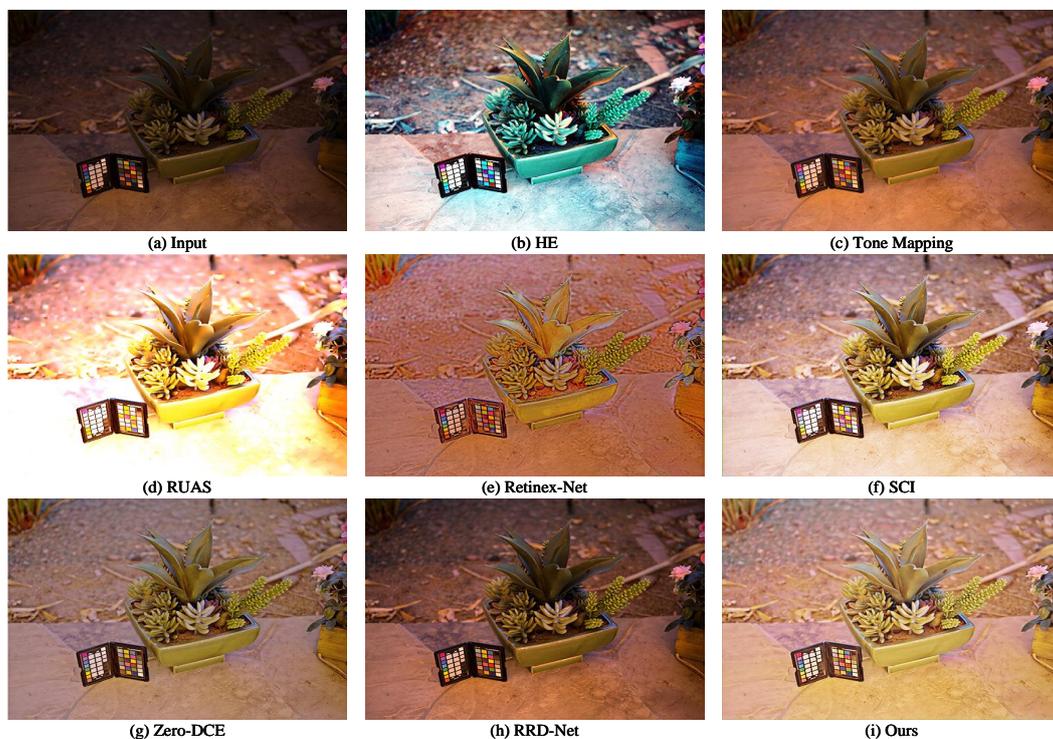


Figure 9. Visual comparisons of different approaches on the LIME benchmark.



Figure 10. Visual comparisons of different approaches on the MEF benchmark.

4.3. Objective Evaluation

We evaluate the results of the proposed method and seven other representative methods on the LOL and VE-LOL paired datasets. Table 1 shows the average MAE, PSNR, SSIM, and LPIPS scores of these two public datasets. Among these evaluation indexes, the higher the PNSR and SSIM values, the better the image quality. On the contrary, the smaller the MAE and LPIPS, the better the image quality. From Table 1, it is evident that our method outperforms other approaches significantly on both test sets, demonstrating the effectiveness of the DBENet framework we proposed.

Table 1. Quantitative comparison on LOL and VE-LOL datasets. The best result is in bold, whereas the second best results are in underline, respectively.

Dataset	Method	MAE↓	PSNR↑	SSIM↑	LPIPS↓
LOL	Input	0.3914	7.7733	0.1952	0.4191
	HE	0.1879	12.918	0.3369	0.4376
	Tone Mapping	0.1517	16.5034	0.5092	0.2312
	RUAS	0.1534	16.4047	0.4997	<u>0.1937</u>
	Retinex-Net	<u>0.1255</u>	<u>16.7740</u>	0.4196	0.3758
	SCI	0.1912	14.784	0.5220	0.2385
	Zero-DCE	0.1860	14.7971	<u>0.5573</u>	0.2368
	RRDNet	0.2739	11.4037	0.4575	0.2480
	Ours	0.1007	19.8625	0.8149	0.1152
VE-LOL (real)	Input	0.3131	9.7168	0.1989	0.3472
	HE	0.4901	13.1314	0.3760	0.4140
	Tone Mapping	<u>0.1298</u>	<u>17.2469</u>	0.5262	0.2349
	RUAS	0.1621	15.3255	<u>0.4878</u>	0.2165
	Retinex-Net	0.1313	16.0971	<u>0.4011</u>	0.4368
	SCI	0.1470	17.3035	0.5336	0.2021
	Zero-DCE	0.1320	17.9992	0.5719	0.2154
	RRDNet	0.2089	13.9818	0.4832	<u>0.1896</u>
	Ours	0.0099	19.8285	0.8437	0.1086

In addition, we also evaluated these datasets using the non-reference image quality evaluator (NIQE), as shown in Table 2. With the exception of Zero-DCE, which had the best score on some datasets, our NIQE scores outperformed most of the other methods. Overall, Tables 1 and 2 provide stronger evidence for the effectiveness and applicability of our proposed method.

Table 2. NIQE scores on low-light image sets (LOL, VE-LOL, LIME, and MEF). The best result is in bold, whereas the second best results are in underline, respectively. Smaller NIQE scores indicate a better quality of perceptual tendency.

Method	LOL	VE-LOL (Real)	LIME	MEF
HE	8.1541	8.7654	6.8883	3.5638
Tone Mapping	7.8310	7.9683	3.9201	3.5254
RUAS	<u>6.3400</u>	<u>6.5330</u>	5.3642	5.4255
Retinex-Net	8.8781	9.4276	4.7669	4.4097
SCI	7.8766	8.0461	4.2064	3.6277
Zero-DCE	7.7925	8.0449	<u>3.9733</u>	3.3023
RRDNet	7.4777	7.7131	4.0689	<u>3.4796</u>
Ours	5.2485	5.0481	4.3475	4.2920

4.4. Ablation Study

We conducted ablation studies on the dual-branch network, and the data results are shown in Table 3. The CNN branch is based on spatially extracting local features from the image, which may overlook global contextual relationships that are crucial for understanding the overall representation. On the other hand, the encoder–decoder branch-based method captures global contextual relationships via skip connections but may overlook local features, which can affect the fusion outcome. We performed experiments on three different methods, including a single branch and a combination of both branches. The experimental results indicate that our proposed dual-branch fusion network outperforms the CNN branch or U-Net branch methods in all metrics. Therefore, combining the capture of global contextual relationships and local features can improve the fusion-enhancement effect for low-light images.

Table 3. Data of ablation experiment.

Methods	PSNR	SSIM
CNN branch	18.6481	0.7128
U-Net branch	17.5846	0.7813
DBENet	19.8625	0.8149

5. Discussion

To shed light on the core mechanisms underpinning our model’s exceptional performance, we introduce DBENet, a deep-learning framework designed explicitly for enhancing and denoising low-light images. Our model adopts a divide-and-conquer strategy, breaking down the intricacies into manageable components for separate handling. Furthermore, we combine the improved gamma correction with deep learning, as illustrated in Figure 11. The regions highlighted within the red boxes demonstrate that our approach avoids excessive amplification of well-exposed parts of the input image. This approach enables us to carefully balance image fidelity while enhancing brightness.

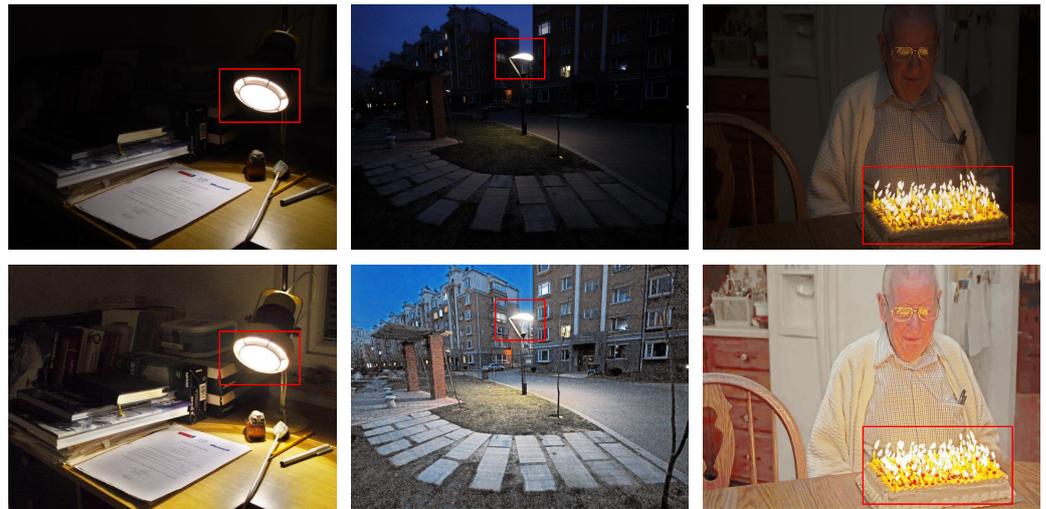


Figure 11. Visual comparison examples of non-uniform illumination images. The top images represent the input, while the bottom images depict the model's output. In particular, within the red rectangles, the light sources are not excessively enhanced.

Moreover, this research opens opportunities for future investigations. These prospects include the reduction in model inference time, enabling the real-time processing of high-resolution visuals, and exploring applications in low-light video enhancement. These endeavors hold significant potential for advancing the frontiers of image and video enhancement across a diverse range of real-world scenarios.

6. Conclusions

We propose an end-to-end dual-branch low-light enhancement architecture network based on the YCbCr color space, inspired by the separation of luminance and chrominance information in YCbCr color space. This network aims to address the issues of brightness distortion, color distortion, and noise pollution in enhanced images caused by the high coupling between brightness and RGB channels in low-light images. The enhancement network adopts a dual-branch structure to enhance the contrast of the luminance channel and suppress the noise in the chrominance channel. The experimental results demonstrate that our proposed method effectively enhances brightness, restores image textures, and produces images with richer details, more realistic colors, and less noise. Compared to classical low-light enhancement algorithms, our approach achieves significant improvements in multiple metrics and multiple datasets, while being more lightweight and faster in processing speed.

Author Contributions: Conceptualization, Y.C., C.W. and W.L.; methodology, C.W.; software, Y.C.; validation, Y.C. and W.H.; formal analysis, Y.C.; investigation, Y.C.; resources, Y.C.; data curation, Y.C.; writing—original draft preparation, Y.C.; writing—review and editing, C.W. and W.L.; visualization, W.H.; supervision, C.W. and W.L.; project administration, C.W.; funding acquisition, C.W. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant No. 62125307 and supported by the Opening Project of Guangdong Provincial Key Lab of Robotics and Intelligent System.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: These data can be found here: LOL <https://daooshee.github.io/BMVC2018website/>, VE-LOL https://flyywh.github.io/IJCV2021LowLight_VELOL/, and LIME and MEF <https://drive.google.com/drive/folders/1lp6m5JE3kf3M66Dicbx5wSnhxt90V4T>, accessed on 10 August 2018, 11 January 2021, and 23 June 2019, respectively.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclatures

DBENet	Dual-Branch Brightness Enhancement Fusion Network
FM	Fusion Module
MAE	Mean Absolute Error
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index Measure
LPIPS	Learned Perceptual Image Patch Similarity
NIQE	Natural Image Quality Evaluator

References

- Meng, L.; Li, H.; Chen, B.C.; Lan, S.; Wu, Z.; Jiang, Y.G.; Lim, S.N. Adavit: Adaptive vision transformers for efficient image recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12309–12318.
- Fang, W.; Wang, L.; Ren, P. Tinier-YOLO: A real-time object detection method for constrained environments. *IEEE Access* **2019**, *8*, 1935–1944. [[CrossRef](#)]
- Strudel, R.; Garcia, R.; Laptev, I.; Schmid, C. Segmenter: Transformer for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 7262–7272.
- Fujiyoshi, H.; Hirakawa, T.; Yamashita, T. Deep learning-based image recognition for autonomous driving. *IATSS Res.* **2019**, *43*, 244–252. [[CrossRef](#)]
- Stark, J.A. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.* **2000**, *9*, 889–896. [[CrossRef](#)]
- Reza, A.M. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *J. Vlsi Signal Process. Syst. Signal, Image Video Technol.* **2004**, *38*, 35–44. [[CrossRef](#)]
- Li, L.; Si, Y.; Jia, Z. Remote sensing image enhancement based on non-local means filter in NSCT domain. *Algorithms* **2017**, *10*, 116. [[CrossRef](#)]
- Lam, E.Y. Combining gray world and retinex theory for automatic white balance in digital photography. In Proceedings of the Ninth International Symposium on Consumer Electronics, Macau SAR, China, 14–16 June 2005; pp. 134–139.
- Xie, C.; Tang, H.; Fei, L.; Zhu, H.; Hu, Y. IRNet: An Improved Zero-Shot Retinex Network for Low-Light Image Enhancement. *Electronics* **2023**, *12*, 3162. [[CrossRef](#)]
- Park, J.S.; Soh, J.W.; Cho, N.I. Generation of high dynamic range illumination from a single image for the enhancement of undesirably illuminated images. *Multimed. Tools Appl.* **2019**, *78*, 20263–20283. [[CrossRef](#)]
- Xu, H.; Ma, J.; Jiang, J.; Guo, X.; Ling, H. U2Fusion: A unified unsupervised image fusion network. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 502–518. [[CrossRef](#)]
- Deng, X.; Zhang, Y.; Xu, M.; Gu, S.; Duan, Y. Deep Coupled Feedback Network for Joint Exposure Fusion and Image Super-Resolution. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* **2021**, *30*, 3098–3112. [[CrossRef](#)]
- Lu, H.; Gong, J.; Liu, Z.; Lan, R.; Pan, X. FDMLNet: A Frequency-Division and Multiscale Learning Network for Enhancing Low-Light Image. *Sensors* **2022**, *22*, 8244. [[CrossRef](#)]
- Guo, X.; Hu, Q. Low-light image enhancement via breaking down the darkness. *Int. J. Comput. Vis.* **2023**, *131*, 48–66. [[CrossRef](#)]
- Zhang, J.; Ji, R.; Wang, J.; Sun, H.; Ju, M. DEGAN: Decompose-Enhance-GAN Network for Simultaneous Low-Light Image Lightening and Denoising. *Electronics* **2023**, *12*, 3038. [[CrossRef](#)]
- Ahn, H.; Keum, B.; Kim, D.; Lee, H.S. Adaptive local tone mapping based on retinex for high dynamic range images. In Proceedings of the 2013 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 11–14 January 2013; pp. 153–156.
- Huang, S.C.; Cheng, F.C.; Chiu, Y.S. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Trans. Image Process.* **2012**, *22*, 1032–1041. [[CrossRef](#)] [[PubMed](#)]
- Ooi, C.H.; Isa, N.A.M. Quadrants dynamic histogram equalization for contrast enhancement. *IEEE Trans. Consum. Electron.* **2010**, *56*, 2552–2559. [[CrossRef](#)]
- Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **2016**, *26*, 982–993. [[CrossRef](#)]
- Jobson, D.J.; Rahman, Z.u.; Woodell, G.A. Properties and performance of a center/surround retinex. *IEEE Trans. Image Process.* **1997**, *6*, 451–462. [[CrossRef](#)]
- Rahman, Z.u.; Jobson, D.J.; Woodell, G.A. Multi-scale retinex for color image enhancement. In Proceedings of the 3rd IEEE International Conference on Image Processing, Austin, TX, USA, 19 September 1996; Volume 3, pp. 1003–1006.
- Jobson, D.J.; Rahman, Z.u.; Woodell, G.A. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.* **1997**, *6*, 965–976. [[CrossRef](#)]

23. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vision Graph. Image Process.* **1987**, *39*, 355–368. [[CrossRef](#)]
24. Land, E.H.; McCann, J.J. Lightness and retinex theory. *Josa* **1971**, *61*, 1–11. [[CrossRef](#)]
25. Lore, K.G.; Akintayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **2017**, *61*, 650–662. [[CrossRef](#)]
26. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep Retinex Decomposition for Low-Light Enhancement. In Proceedings of the British Machine Vision Conference, British Machine Vision Association, Newcastle, UK, 3–6 September 2018.
27. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [[CrossRef](#)] [[PubMed](#)]
28. Lv, F.; Lu, F.; Wu, J.; Lim, C. MBLLEN: Low-Light Image/Video Enhancement Using CNNs. In Proceedings of the BMVC, Newcastle, UK, 3–6 September 2018; Volume 220, p. 4.
29. Zheng, C.; Shi, D.; Shi, W. Adaptive Unfolding Total Variation Network for Low-Light Image Enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; pp. 4439–4448.
30. Wu, W.; Weng, J.; Zhang, P.; Wang, X.; Yang, W.; Jiang, J. URetinex-Net: Retinex-Based Deep Unfolding Network for Low-Light Image Enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 5901–5910.
31. Chenxi Wang Hongujun Wu and Zhi Jin. FourLLIE: Boosting Low-Light Image Enhancement by Fourier Frequency Information. In Proceedings of the ACM MM, Thessaloniki, Greece, 12–15 June 2023.
32. Liu, R.; Ma, L.; Zhang, J.; Fan, X.; Luo, Z. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10561–10570.
33. Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1780–1789.
34. Li, C.; Guo, C.; Feng, R.; Zhou, S.; Loy, C.C. CuDi: Curve Distillation for Efficient and Controllable Exposure Adjustment. *arXiv* **2022**, arXiv:2207.14273.
35. Zhu, A.; Zhang, L.; Shen, Y.; Ma, Y.; Zhao, S.; Zhou, Y. Zero-shot restoration of underexposed images via robust retinex decomposition. In Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME), Virtual, 6–10 July 2020; pp. 1–6.
36. Ma, L.; Ma, T.; Liu, R.; Fan, X.; Luo, Z. Toward fast, flexible, and robust low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5637–5646.
37. Hue Nguyen and Diep Tran and Khoi Nguyen and Rang Nguyen. PSENet: Progressive Self-Enhancement Network for Unsupervised Extreme-Light Image Enhancement. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2–7 January 2023.
38. Ko, K.; Kim, C.-S. IceNet for interactive contrast enhancement. *IEEE Access* **2021**, *9*, 168342–168354. [[CrossRef](#)]
39. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015*; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
40. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
41. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
42. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
43. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.
44. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.