

Article

Adaptive Dual Aggregation Network with Normalizing Flows for Low-Light Image Enhancement

Hua Wang^{1,2,*} , Jianzhong Cao¹ and Jijiang Huang¹

¹ Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China; cjz@opt.ac.cn (J.C.); huangjijiang@opt.ac.cn (J.H.)

² University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: wanghua@opt.ac.cn

Abstract: Low-light image enhancement (LLIE) aims to improve the visual quality of images taken under complex low-light conditions. Recent works focus on carefully designing Retinex-based methods or end-to-end networks based on deep learning for LLIE. However, these works usually utilize pixel-level error functions to optimize models and have difficulty effectively modeling the real visual errors between the enhanced images and the normally exposed images. In this paper, we propose an adaptive dual aggregation network with normalizing flows (ADANF) for LLIE. First, an adaptive dual aggregation encoder is built to fully explore the global properties and local details of the low-light images for extracting illumination-robust features. Next, a reversible normalizing flow decoder is utilized to model real visual errors between enhanced and normally exposed images by mapping images into underlying data distributions. Finally, to further improve the quality of the enhanced images, a gated multi-scale information transmitting module is leveraged to introduce the multi-scale information from the adaptive dual aggregation encoder into the normalizing flow decoder. Extensive experiments on paired and unpaired datasets have verified the effectiveness of the proposed ADANF.

Keywords: low-light image enhancement; normalizing flow; adaptive dual aggregation; deep learning



Citation: Wang, H.; Cao, J.; Huang, J. Adaptive Dual Aggregation Network with Normalizing Flows for Low-Light Image Enhancement. *Entropy* **2024**, *26*, 184. <https://doi.org/10.3390/e26030184>

Academic Editors: Oleg Sergiyenko, Wendy Flores-Fuentes, Julio Cesar Rodríguez-Quinones and Jesús Elías Miranda-Vega

Received: 8 January 2024

Revised: 18 February 2024

Accepted: 20 February 2024

Published: 22 February 2024



Copyright: © 2024 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

Insufficient light in complex imaging environments can lead to dark brightness, low contrast, high noise, and poor details in captured images [1,2]. Low-light image enhancement (LLIE) aims to solve the problems of insufficient visibility and low contrast in low-light images while restoring noise, structures, color distortion, etc. [3]. Low-light image enhancement can effectively improve the performance of methods such as object detection and scene understanding at night or in low-light conditions [4].

Over the past decades, many low-light image enhancement methods have been proposed [5,6]. Previous methods are usually based on hand-designed features and processing steps such as histogram equalization [7,8] and gamma transformation [9]. These methods are simple and fast, but they usually amplify noise while enhancing the image and often cannot restore the color and details of low-light images well [10]. The widely popular Retinex theory [11] provides an intuitive and easy-to-understand framework for LLIE by decomposing the image into reflection and illumination components [12,13]. However, for complex illumination properties in practice, it is challenging to design priors and regularizations that are always valid for accurate decomposition of the reflection and illumination components [14,15]. Improper decomposition can lead to unrealistic details, undesirable artifacts, and color distortion in enhanced images [16].

Inspired by the successful application of deep learning in object recognition, detection, etc. [17,18], researchers have focused on building various deep learning frameworks suitable for the LLIE task [6,19,20]. In addition, the development of paired datasets [21,22] has

indeed been a critical step in enabling the application of deep learning to the LLIE task. Recent LLIE methods based on deep learning can be roughly divided into deep-Retinex-based methods and end-to-end methods [5,23,24].

Deep-Retinex-based methods are also based on the human visual system's Retinex theory and use neural networks to simulate the process of separating the reflectance and illumination components [25]. These methods aim to combine the advantages of both Retinex theory and deep learning, enabling an interpretable low-light image enhancement paradigm [26,27]. Under low-light conditions, the boundary between the reflectance component and the illuminance component can become blurred, making it more difficult to accurately separate them. Even when using deep learning models for Retinex decomposition, there is still the problem of being unable to accurately separate the reflectance and illumination components, which may lead to noise amplification and image stylization in the enhanced results [28]. In order to solve these problems, researchers are constantly exploring ways to improve deep learning models, such as by using more complex network structures or introducing regularization techniques to improve model performance [10,29].

End-to-end methods typically use deep neural networks to directly learn the non-linear relationships between low-light images and their corresponding normally exposed images [30,31]. By removing the need for explicit separation of the reflectance and illumination components, end-to-end methods focus on designing a variety of novel neural network structures for LLIE [32]. End-to-end methods have the advantage of being less dependent on physical models and can directly learn the desired mapping between low-light and normal-light images [33]. However, they may not be as interpretable as deep-Retinex-based methods due to their black-box nature.

Recent methods based on deep learning have made good progress in LLIE. However, these methods generally use pixel-level error functions, such as L1 or L2 norm, as the objective function of deep networks for training [5,10]. Pixel-level error functions cannot measure the real visual errors between enhanced images and normally exposed images such as complex structures and textures [34,35]. And pixel-level error functions have difficulty providing effective regularization for local structures in various complex backgrounds.

To alleviate the above problem, we propose an adaptive dual aggregation network with normalizing flows (ADANF) for low-light image enhancement. Different from previous methods that use pixel-level error functions to measure the difference between enhanced and normally exposed images in the image domain, we adopt a normalizing flow framework to map enhanced and normally exposed images to the underlying data distribution, which can effectively express the structural details of complex images [36]. In addition, we use the errors between the data distributions for enhanced and normally exposed images as the objective function to effectively measure the visual distance.

In the proposed ADANF, an adaptive dual aggregation encoder is firstly exploited to extract illumination-robust features by fully exploring the global properties and local details of the low-light images. Next, a reversible normalizing flow decoder is leveraged to recover normally exposed images from the illumination-robust features. Here, we exploit the inverse process capabilities of the normalized stream decoder to reconstruct brighter, more detailed images. Finally, to further improve the quality of image enhancement, a gated multi-scale information transmitting module is designed to introduce the multi-scale features from the adaptive dual aggregation encoder into the normalizing flow decoder. Extensive experiments on paired and unpaired datasets verify the effectiveness of the proposed ADANF.

The contributions of this paper mainly include:

- An adaptive dual aggregation encoder is leveraged to fully capture the global properties and local details of low-light images for extracting illumination-robust features from low-light images.
- To measure real visual errors between enhanced and normally exposed images, a reversible normalizing flow decoder is used to map enhanced and normally exposed

images to potential distributions, and the difference between the distributions is used as the objective function for training.

- A gated multi-scale information transmitting module is designed to introduce the multi-scale features from the adaptive dual aggregation encoder into the normalizing flow decoder to further improve the quality of enhanced images.

The rest of the manuscript is organized as follows. Recent related works are introduced in Section 2. Section 3 gives the details of the proposed ADANF. Section 4 reports experimental results. Finally, the conclusion is provided in Section 5.

2. Related Work

2.1. Traditional Methods

Previous methods usually study hand-designed features for LLIE. Histogram equalization is one of the most classic low-light image enhancement methods [37]. Reza [38] designed a block-based histogram equalization method to model lighting changes in local areas. Lee et al. [39] calculated the 2D histogram by considering the relationship between neighboring pixels within local regions. They utilized the layered difference approach for enhancing contrast. In addition, some researchers attempted to combine image quality assessment with histogram equalization to improve performance. Gu et al. [40] used subjective and objective evaluation guidance to improve the histogram to correct image brightness and contrast to the level of normal exposure.

Retinex theory is also very popular in low-light image enhancement, and researchers have carefully designed many decomposition methods based on the Retinex theory. Kimmel et al. [41] proposed to introduce the lighting component gradient into a variational framework for LLIE. Ren et al. [12] designed a low-rank prior regularized Retinex decomposition model to alleviate the noise amplification problem. Gu et al. [13] proposed a fractional-order variational structure that regularizes both the reflectance and illumination components. Liang et al. [42] combined nonlinear diffusion techniques and Retinex decomposition to estimate lighting components to improve estimation results. These methods are sensitive to illumination changes. In low-light environments, illumination changes may lead to inaccurate feature extraction and affect the enhancement effect.

2.2. Deep-Learning-Based Methods

Recent methods mainly study the design of deep learning frameworks for LLIE, including deep-Retinex-based methods and end-to-end methods [5,10]. Deep-Retinex-based methods combine the advantages of Retinex theory and deep learning to provide an interpretable solution for low-light image enhancement. Wei et al. [21] proposed a Retinex-Net including Decom-Net and Enhance-Net. Decom-Net is responsible for decomposing the input low-light image into reflection and illumination parts, while Enhance-Net is responsible for enhancing the illumination part to obtain normally exposed images. Zhang et al. [43] proposed a KinD network to utilize images under different exposure conditions for training. Fan et al. [29] introduced a semantic segmentation sub-network into the Retinex model to use semantic priors to guide image enhancement. Liu et al. [27] employed unrolling and adjustment to exploit global and local brightness of images for LLIE.

End-to-end methods focus on carefully designing different networks to learn the mapping between low-light and normally exposed images. Lore et al. [20] designed the first deep network LLNet for LLIE, which is a sparse denoising autoencoder structure. Yang et al. [33] proposed to exploit a transformer-based network to extract the global information of low-light images. Ren et al. [44] utilized an encoder–decoder network to extract global content and a recurrent neural network to preserve edge details. Xu et al. [45] proposed a frequency-based model that uses low-frequency layers to restore content and high-frequency layers to restore image details. Xu et al. [31] considered that the information amounts in different areas are different and designed a signal-to-noise-ratio-aware transformer for LLIE. However, recent deep-learning-based methods usually employ pixel-level

L1 or L2 norm as the objective function to optimize deep networks, which cannot effectively measure the real visual errors between the enhanced image and the normal exposure image.

3. Methods

LLIE aims at generating the normally exposed image $X_n \in \mathbb{R}^{H \times W \times 3}$ from a low-light image $X_l \in \mathbb{R}^{H \times W \times 3}$, where W and H represent the width and height, respectively. Previous methods focus on studying different networks, directly utilizing MSE [20], L1 [46], or color loss [47] as objective functions to perform supervised training under paired training samples $\{X_l, X_{gt}\}$, where $X_{gt} \in \mathbb{R}^{H \times W \times 3}$ is the ground truth normally exposed image. However, there are two problems with previous methods. First, it is difficult for these methods to fully adaptively utilize the global and local information of the image X_l to improve visual effect and suppress noise. On the other hand, the loss functions of these methods focus on pixel level or local errors, and it is difficult to fully utilize the visual properties to measure the real visual errors between the generated image X_n and the ground truth X_{gt} [35].

To alleviate the above two problems, an adaptive dual aggregation network with normalizing flows (ADANF) is proposed for LLIE. The overall structures of the ADANF are shown in Figure 1. First, an adaptive dual aggregation encoder is employed to fully exploit the global properties and local details of the image X_l to extract illumination-robust features. Then, an invertible normalizing flow decoder is used to recover the normally exposed image X_n from the illumination-robust features. Finally, a gated multi-scale information transmitting module is designed to introduce the multi-scale features of the adaptive dual aggregation encoder into the normalizing flow decoder to further improve the quality of image enhancement.

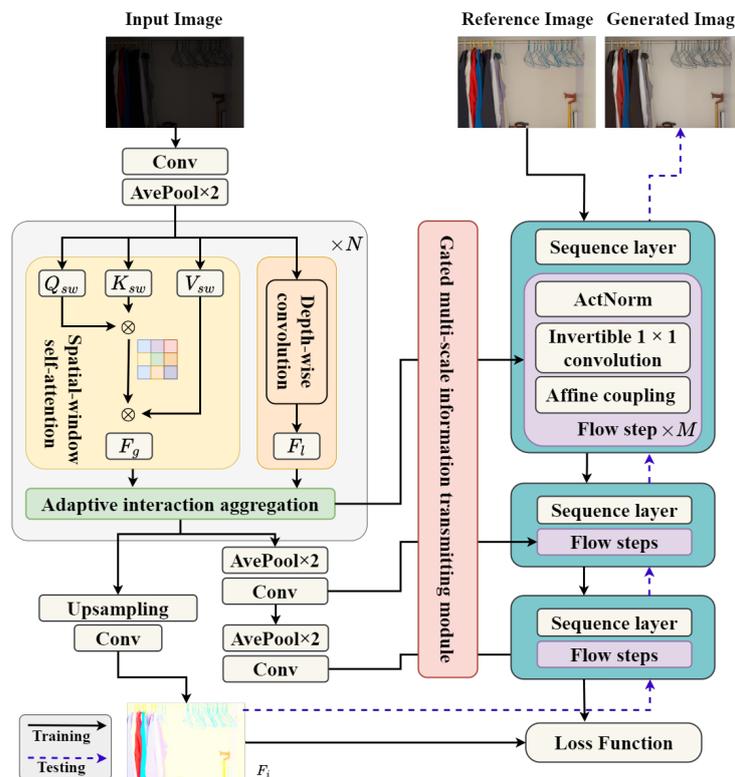


Figure 1. Detailed structures of ADANF. In the testing phase, a low-light image is first fed to the adaptive dual aggregation encoder to fully exploit the global properties and local details for extracting illumination-robust features. Then, a gated multi-scale information transmitting module is designed to introduce the multi-scale features of the adaptive dual aggregation encoder into the normalizing flow decoder. Finally, an invertible normalizing flow decoder is used to recover the normally exposed image from the illumination-robust features.

3.1. Adaptive Dual Aggregation Encoder

3.1.1. Preprocessing

Low-light images often have local or global dark areas, resulting in poor contrast and unclear detail. In addition, insufficient light may also cause problems such as noise and artifacts. If the original low-light images are input directly into the model, the model may have difficulty distinguishing low-contrast areas and noisy areas. By performing histogram equalization on X_l , we can redistribute the pixel intensities of an image X_l so that they occupy the entire possible intensity range. The histogram-equalized image $h(X_l) \in \mathbb{R}^{H \times W \times 3}$ will have higher contrast and the model can more easily identify and perceive different areas in the image. In addition, we use color map $c(X_l) \in \mathbb{R}^{H \times W \times 3}$ to enhance the contrast and visibility of low-light images X_l , highlighting details in dark areas, where $c(X_l) = X_l / \text{mean}_p(X_l)$, and $\text{mean}_p(X_l)$ represents the calculation of the mean value of each pixel in X_l . Finally, we use the gradient map $g(X_l) \in \mathbb{R}^{H \times W \times 3}$ to explicitly capture the noisy areas in low-light images X_l , where $g(X_l) = \max(|\nabla_x(c(X_l))|, |\nabla_y(c(X_l))|)$, ∇_x , and ∇_y are the gradients in the x and y directions, respectively. To improve the model's sensitivity to noisy areas in low-contrast and dark areas, $h(X_l)$, $c(X_l)$, $g(X_l)$, and X_l are stacked by channel as the input $X_{in} = [h(X_l), c(X_l), g(X_l), X_l]$ of the subsequent network.

3.1.2. Global–Local Adaptive Aggregation Module

In the adaptive dual aggregation encoder, two 3×3 convolutions are first used to transform the image X_{in} into the feature space to obtain the shallow feature $F_s \in \mathbb{R}^{H \times W \times C_s}$, where C_s is the channel number. Then, global–local adaptive aggregation blocks are used to extract illumination-robust feature $F_i \in \mathbb{R}^{H \times W \times C_i}$. The global–local adaptive aggregation block is the key module of the adaptive dual aggregation encoder, and we take a global–local adaptive aggregation block as an example to introduce its details.

First, spatial-window self-attention [48,49] is utilized to explore the global information of the image. We generate query features $Q \in \mathbb{R}^{H \times W \times C_i}$, key features $K \in \mathbb{R}^{H \times W \times C_i}$, and value features $V \in \mathbb{R}^{H \times W \times C_i}$ from the shallow feature F_s by using convolutions.

$$Q = W_Q F_i, K = W_K F_i, V = W_V F_i, \quad (1)$$

where $W_Q, W_K, W_V \in \mathbb{R}^{1 \times 1 \times C_i}$ are the weights of a 1×1 convolution, and biases are omitted. Since performing self-attention directly on the global image will introduce a huge amount of calculation, we follow SwinTransformer [50] to perform the spatial-window self-attention to reduce the amount of calculation. Q, K , and V are divided into non-overlapping spatial windows Q_{sw}^j, K_{sw}^j , and $V_{sw}^j \in \mathbb{R}^{H_{sw} \times W_{sw} \times C_i}$, respectively. $H_{sw} \times W_{sw}$ is the size of the spatial window. We can calculate the features of each spatial window using Equation (2).

$$F_g^j = \text{softmax}(Q_{sw}^j (K_{sw}^j)^T / \sqrt{C_i} + P^j) V_{sw}^j, \quad (2)$$

where P^j is relative position encoding of the j -th spatial window. The outputs of spatial-window self-attention are $F_g = [F_g^1, F_g^2, \dots, F_g^n]$, where $n = (H/H_{sw})^2$ is the number of spatial windows and $F_g \in \mathbb{R}^{H \times W \times C_i}$. In addition, shift window operations [50] are utilized to extract the global spatial feature of the image.

Second, to capture details and textures in images of LLIE, a local branch uses depth-wise convolution (DWC) operations to extract local features $F_l = \text{DWC}(V) \in \mathbb{R}^{H \times W \times C_i}$ from the value features V from Equation (1).

Third, to fully utilize the global and local information of the image X_l to generate illumination-robust features, an adaptive interaction aggregation (AIA) module is designed. Since F_g is the global information of the image and F_l is the local features of the image, F_g and F_l are misaligned features. In this case, simple feature weighted combination or concatenation operations cannot fully integrate global and local information. In the AIA module, we first use the information of local features F_l to refine the texture detail

information of global features F_g by exploiting the attention mechanism. The spatial attention map $S(F_l) \in \mathbb{R}^{H \times W \times 1}$ of the local features is calculated as

$$S(F_l) = \varphi(W_{sa2}\sigma(W_{sa1}F_l)), \quad (3)$$

where φ is the sigmoid activation, σ is the RELU activation, $W_{sa1}, W_{sa2} \in \mathbb{R}^{1 \times 1 \times C_i}$ are weights of the 1×1 convolutions, W_{sa1} contains C_i kernels, and W_{sa2} contains a kernel. Then, we can obtain the refined global feature $\hat{F}_g = F_g \odot S(F_l)$, where \odot is the Hadamard product, $\hat{F}_g \in \mathbb{R}^{H \times W \times C_i}$. Then, the AIA module utilizes the rich channel information of global features F_g to suppress redundant channels of local features F_l . The channel attention map $C(F_g) \in \mathbb{R}^{1 \times 1 \times C_i}$ of F_g is

$$C(F_g) = \varphi(W_{ca2}\sigma(W_{ca1}\text{GAP}(F_g))), \quad (4)$$

Then, we can obtain the refined local feature $\hat{F}_l = F_l \odot C(F_g)$, $\hat{F}_l \in \mathbb{R}^{H \times W \times C_i}$.

Finally, the refined global and local features \hat{F}_g and \hat{F}_l are aggregated by element-wise addition as the output. Multiple global–local adaptive aggregation blocks are repeated to generate illumination-robust feature F_i .

3.2. Normalizing Flow Decoder

During real imaging, changes in lighting conditions (e.g., different time, weather, or light sources) can cause even the same scene to look completely different in low-light images. That is, a normally exposed image will correspond to many different low-light images. A good LLIE method should be able to adapt to changes in lighting conditions. In this paper, we propose to exploit a normalizing flow decoder to recover normally exposed images from illumination-robust feature F_i .

In the proposed ADANF, the normalizing flow decoder is an invertible network, whose purpose is to learn a one-to-many mapping relationship for LLIE. In the training phase, the normalizing flow decoder aims to learn the mapping of normally exposed images to the feature F_i of low-light images [51,52]. The normalizing flow network can adapt to various characteristics of the same scene under different lighting conditions. During the testing phase, the inverse of the learned mapping can be exploited to generate normally exposed images from low-light image features F_i .

The structures of the normalizing flow decoder has three levels, with a squeeze layer and 12 flow steps at each level. A squeeze layer is a type of layer that reduces the spatial resolution of the input data, which can help with reducing the computational complexity of the network. The flow steps are the main part of the invertible network, where the invertible mapping from normally exposed images to the feature of low-light images is learned.

As shown in Figure 1, a flow step is composed of an activation normalization (ActNorm) layer, an invertible 1×1 convolution, and an affine coupling component. The ActNorm layer is similar to batch normalization, using the scale $\mu \in \mathbb{R}^{1 \times 1 \times C_i}$ and bias $\sigma \in \mathbb{R}^{1 \times 1 \times C_i}$ parameters of each channel of the input data to perform a transformation $Y_i = \frac{F_i - \mu}{\sigma} \in \mathbb{R}^{H \times W \times C_i}$ as preprocessing, whose purpose is to make the input data F_i have zero mean and unit variance. The scale μ and bias σ parameters of the ActNorm layer are learnable and initialized using the mean and variance of batch features.

After the ActNorm layer, an invertible 1×1 convolution is used to increase the information interaction between the feature channels of Y_i to obtain $\tilde{Y}_i \in \mathbb{R}^{H \times W \times C_i}$. In invertible 1×1 convolutions, given the output data and convolution kernel, we can accurately recover the original input data. In this way, we can reconstruct the normally exposed image from the feature F_i based on the inverse of the learned mapping. In order to make the traditional 1×1 convolution invertible, we need to set its weight matrix to a random orthogonal matrix [51].

The affine coupling component is a special reversible transformation component that can effectively map input data to different feature spaces. It transforms existing channels through multiplication and addition operations and can effectively facilitate

the normalizing flow decoder to learn the mapping from normally exposed images to the feature F_i during the training phase. In the affine coupling component, a split operation is first utilized to divide the input data \tilde{Y}_i into two data, $\tilde{Y}_i^1 \in \mathbb{R}^{H \times W \times C_i/2}$ and $\tilde{Y}_i^2 \in \mathbb{R}^{H \times W \times C_i/2}$ along the channel dimension. Then, we perform an identity transformation on \tilde{Y}_i^1 to obtain $H_1 = \tilde{Y}_i^1$ and perform an affine transformation on \tilde{Y}_i^2 to obtain H_2 ,

$$H_2 = \exp\left(\text{NN}_s\left(\tilde{Y}_i^1\right)\right) \odot \tilde{Y}_i^2 + \text{NN}_b\left(\tilde{Y}_i^1\right), \tag{5}$$

where $\text{NN}_s(\tilde{Y}_i^1)$ and $\text{NN}_b(\tilde{Y}_i^1)$ are shadow three-layer convolutional neural networks to learn the scale and bias from \tilde{Y}_i^1 for affine transformation. Next, H_1 and H_2 are concatenated by channel and input into invertible 1×1 convolutions for information interaction among channels. Similar to recent methods [51,52], the flow step is repeated 12 times to learn the mapping.

3.3. Mapping Learning Aided by Multi-Scale Features

Due to complex low-light conditions, the detailed information of the image at different scales will be lost or obscured, or the areas at different scales will be too dark or too bright [53]. The multi-scale information of the image is important for LLIE, but the above normalizing flow decoder cannot effectively utilize its multi-scale information. In the proposed ADANF, a gated multi-scale information transmitting module is used to introduce the multi-scale features of the adaptive dual aggregation encoder into the normalizing flow decoder to further improve the quality of image enhancement.

Detailed structures of the gated multi-scale information transmitting module are shown in Figure 2. Three dilated convolutions with different dilation rates (e.g., [1,2,4]) are firstly used in parallel to extract the features of different scales [54]. Then, these features are concatenated and fed to 1×1 convolutions to generate multi-scale features $F_{ms} \in \mathbb{R}^{H \times W \times C_i}$. Next, Global Average-Pooling (GAP) and Global Max-Pooling (GMP) operations in the channel dimension are utilized to extract spatial information $\text{GAP}(F_{ms})$ and $\text{GMP}(F_{ms})$. Convolution operations with sigmoid activation are used to generate attention weights from F_{ms} to control the multi-scale information passed to the normalizing flow decoder. The output of the gated multi-scale information transmitting module is the gated multi-scale features $\bar{F}_{ms} \in \mathbb{R}^{H \times W \times C_i}$,

$$\bar{F}_{ms} = \sigma(\text{Conv}([\text{GAP}(F_{ms}), \text{GMP}(F_{ms})])) \odot F_{ms}. \tag{6}$$

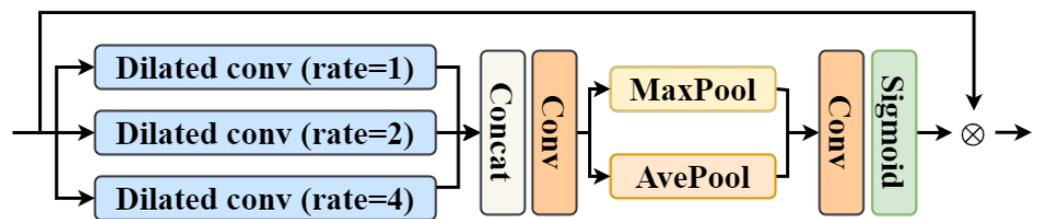


Figure 2. Detailed structures of the gated multi-scale information transmitting module.

3.4. Loss Function

In the ADANF, we use the normalizing flow decoder to capture the conditional distribution $P_{\text{NFD}}(X_{gt}|X_l, \theta)$ of a normally exposed image X_{gt} under its low-light image condition X_l , where θ represents the parameters of the normalizing flow decoder. Since the normalizing flow decoder is an invertible network, it can map a normally exposed image X_{gt} to a latent variable $z = \text{NFD}_\theta(X_{gt}; X_l)$ under the low-light image condition X_l and can also reversibly map the latent variable z to the normally exposed image $X_{gt} = \text{NFD}_\theta^{-1}(z; X_l)$. In ADANF, the latent variable z refers to the illumination-robust feature F_i . Similar to recent work [36], the latent variable z can be assumed to follow a Gaussian distribution $P_z(z)$.

According to the change-of-variables theorem, the conditional distribution $P_{\text{NFD}}(X_{gt}|X_l, \theta)$ can be calculated as:

$$\begin{aligned} P_{\text{NFD}}(X_{gt}|X_l, \theta) &= P_z(z) \left| \det \frac{\partial z}{\partial X_{gt}} \right| \\ &= P_z(\text{NFD}_\theta(X_{gt}; X_l)) \left| \det \frac{\partial \text{NFD}_\theta(X_{gt}; X_l)}{\partial X_{gt}} \right|. \end{aligned} \quad (7)$$

The normalizing flow decoder NFD_θ is sequentially composed of N invertible layers $h^{n+1} = \text{NFD}_\theta^n(h^n; \text{ADAE}^n(X_l))$, where NFD_θ^n is the n -th layer, $n = 0, 1, \dots, N-1$, $h^0 = X_{gt}$, and $h^N = z$. $\text{ADAE}^n(X_l)$ is the latent image features from the adaptive dual aggregation encoder.

According to Equation (7), we can use the negative log-likelihood as a loss function \mathcal{L} to optimize the parameters of the proposed ADANF. By using the chain rule, \mathcal{L} is formulated as:

$$\begin{aligned} \mathcal{L} &= -\log P_z(\text{NFD}_\theta(X_{gt}; X_l)) \\ &\quad - \sum_{n=0}^{N-1} \log \left| \det \frac{\partial \text{NFD}_\theta^n(h^n; \text{ADAE}^n(X_l))}{\partial h^n} \right|. \end{aligned} \quad (8)$$

Since $P_z(z)$ is assumed to follow a Gaussian distribution, $P_z(\text{NFD}_\theta(X_{gt}; X_l))$ can be calculated as:

$$P_z(\text{NFD}_\theta(X_{gt}; X_l)) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z - \text{ADAE}(X_l))^2}{2}\right). \quad (9)$$

In the testing phase, low-light images are input to the adaptive dual aggregation encoder to obtain illumination-robust features, and then these features are input to the normalizing flow decoder through inverse mapping NFD_θ^{-1} to generate normally exposed images.

4. Experiments

4.1. Datasets and Evaluation Metrics

Paired datasets. LOLv1 [21] is one of the most commonly used datasets in LLIE. This dataset is collected from real scenes and contains 500 pairs of low-light and normally exposed images under different lighting conditions. Among them, 485 pairs of images are used for training and 15 pairs of images are used for testing.

LOLv2 [22] contains two subsets, namely LOLv2-real and LOLv2-synthetic. LOLv2-real contains image pairs of different brightness in real scenes obtained by adjusting exposure time and ISO settings. These image pairs are intended to study illumination changes in real application scenarios. Specifically, LOLv2-real contains 689 image pairs for training and 100 image pairs for testing. LOLv2-synthetic synthesizes low-light images from RAW images by analyzing the lighting distribution of low-light images. It contains 1000 image pairs, of which 900 pairs are used for training and 100 pairs are used for testing.

Unpaired datasets. The DICM [55], LIME [3], MEF [56], NPE [57], and VV [46] (<https://sites.google.com/site/vonikakis/datasets>, accessed on 4 September 2023) datasets are real captured images and do not contain normally exposed images as reference images. Therefore, these datasets cannot be used for training. We tested the performance of the proposed ADANF on these several datasets.

Evaluation metrics. For paired datasets like LOL and LOL-v2, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [58] are used as evaluation metrics. PSNR measures the peak signal-to-noise ratio between the original image and the enhanced image, while SSIM takes into account the structural and textural information of the image. In addition, learning perceptual image patch similarity (LPIPS) [59] is also used as an evaluation index, which uses deep features to measure the perceptual similarity of images.

This indicator is learned through deep learning methods. Compared with PSNR and SSIM, LPIPS can more truly reflect the human eye's perception of image quality.

For unpaired datasets such as DICM, LIME, MEF, NPE, and VV, direct evaluation using PSNR, SSIM, or LPIPS is not possible because there are no paired normally exposed images. We use the model parameters trained on LOLv2-synthetic to directly infer the enhanced image. In this case, the Natural Image Quality Evaluator (NIQE) is employed to evaluate the results. For PSNR and SSIM, the larger the value, the better the enhancement quality. For LPIPS and NIQE, the smaller the value, the better the enhancement quality.

4.2. Implement Details

In the proposed ADANF, the number of global-local adaptive aggregation modules in the adaptive dual aggregation encoder is 24, and the normalizing flow decoder has three levels with a squeeze layer and 12 flow steps at each level. The batch size on the LOLv1, LOLv2-real, and LOLv2-synthetic dataset is 8. We train the ADANF for 40,000 iterations using the Adam optimizer with the initial learning rate set to 0.0005 and multiplying the learning rate by 0.5 at 20,000, 30,000, 36,000, and 38,000 iterations. The input image size is set to 160×160 . For unpaired data, we use the parameters trained on LOLv2-synthetic to perform inference and obtain results. All experiments are completed on a dual-card NVIDIA RTX 4090 server.

4.3. Comparisons with the State-of-the-Art Methods on Paired Datasets

In this section, to demonstrate the effectiveness of the proposed ADANF, we compare the ADANF with the state-of-the-art low-light image enhancement methods, such as LIME [3], Zero-DCE [47], RetinexNet [21], DRBN [60], KinD [43], KinD++ [61], EnlightenGAN [62], MIRNet [63], LLFlow [35], LLFormer [16].

The quantitative results of the proposed ADANF and comparison methods on the paired LOLv1, LOLv2-real, and LOLv2-synthetic datasets are reported in Tables 1–3, respectively. Low-light images often suffer from color distortion and low contrast, which make it difficult to extract effective features. Previous methods LIME [3], RetinexNet [21], and KinD [43] usually use classic structures when extracting low-light image features, which makes it difficult to effectively model their complex distribution when processing images under different low-light conditions, thus affecting performance. From Tables 1–3, we can see that our ADANF can be significantly improved under PSNR, SSIM, and LPIPS. It is worth noting that our ADANF has a greater improvement in the PSNR index, indicating that our proposed ADANF can obtain higher quality enhanced normally exposed images. Compared with the recent method LLFormer [16], the PSNR of our ADANF on the LOLv1, LOLv2-real, and LOLv2-synthetic datasets has increased by 0.91%, 1.81%, and 0.66%, respectively. This may be due to the fact that the adaptive dual aggregation encoder can effectively extract the global properties and local details from the low-light images. The proposed gated multi-scale information transmitting module can effectively transfer the latent features from the input image to the normalizing flow decoder so that the enhanced image has a more stable quality. The normalizing flow decoder can effectively model the distribution of normally exposed images to reconstruct high-quality images from illumination-robust features.

Table 1. Quantitative results of the proposed ADANF and the state-of-the-art methods for the LOLv1 datasets. \uparrow/\downarrow means that the larger/smaller the index value, the better/lower the quality. GFLOPs represents the Giga Floating Point Operations. Params represents the number of weight parameters.

| Method | LOLv1 | | | Complexity | |
|-----------------|-----------------|-----------------|--------------------|------------|----------|
| | PSNR \uparrow | SSIM \uparrow | LPIPS \downarrow | GFLOPs | Params/M |
| LIME [3] | 16.76 | 0.560 | 0.350 | - | - |
| Zero-DCE [47] | 14.86 | 0.562 | 0.335 | - | 0.33 |
| RetinexNet [21] | 16.77 | 0.462 | 0.474 | 587.47 | 0.84 |

Table 1. Cont.

| Method | LOLv1 | | | Complexity | |
|----------------------|-------|-------|--------|------------|----------|
| | PSNR↑ | SSIM↑ | LPIPS↓ | GFLOPs | Params/M |
| DRBN [60] | 19.86 | 0.834 | 0.155 | 48.61 | 5.27 |
| KinD [43] | 20.87 | 0.799 | 0.207 | 34.99 | 8.02 |
| KinD++ [61] | 21.30 | 0.823 | 0.175 | - | 9.63 |
| EnlightenGAN [62] | 17.48 | 0.652 | 0.322 | 61.01 | 114.35 |
| MIRNet [63] | 24.14 | 0.842 | 0.131 | 785 | 31.76 |
| LLFlow [35] | 25.13 | 0.872 | 0.117 | - | 37.68 |
| LLFormer [16] | 25.76 | 0.823 | 0.167 | - | 24.55 |
| Diff-Retinex [64] | 21.98 | 0.863 | 0.048 | - | - |
| Transformer-GAN [65] | 23.50 | 0.851 | - | - | - |
| Restormer [66] | 22.43 | 0.823 | - | 144.25 | 26.13 |
| SNR-Aware [31] | 26.72 | 0.851 | 0.152 | 26.35 | 4.01 |
| ADANF (ours) | 26.67 | 0.873 | 0.120 | 252.39 | 117.59 |

Table 2. Quantitative results of the proposed ADANF and the state-of-the-art methods for the LOLv2-real datasets. ↑/↓ means that the larger/smaller the index value, the better/lower the quality. GFLOPs represents the Giga Floating Point Operations. Params represents the number of weight parameters.

| Method | LOLv2-Real | | | Complexity | |
|-------------------|------------|-------|--------|------------|--------|
| | PSNR↑ | SSIM↑ | LPIPS↓ | GFLOPs | Params |
| LIME [3] | 15.24 | 0.470 | 0.415 | - | - |
| Zero-DCE [47] | 18.06 | 0.580 | 0.313 | - | 0.33 |
| RetinexNet [21] | 18.37 | 0.723 | 0.365 | 587.47 | 0.84 |
| DRBN [60] | 20.13 | 0.830 | 0.147 | 48.61 | 5.27 |
| KinD [43] | 17.54 | 0.669 | 0.375 | 34.99 | 8.02 |
| KinD++ [61] | 19.09 | 0.817 | 0.180 | - | 9.63 |
| EnlightenGAN [62] | 18.64 | 0.677 | 0.309 | 61.01 | 114.35 |
| MIRNet [63] | 20.36 | 0.782 | 0.317 | 785 | 31.76 |
| LLFlow [35] | 26.20 | 0.888 | 0.137 | - | 37.68 |
| LLFormer [16] | 26.20 | 0.819 | 0.209 | - | 24.55 |
| Restormer [66] | 19.94 | 0.827 | - | 144.25 | 26.13 |
| SNR-Aware [31] | 27.21 | 0.871 | 0.157 | 26.35 | 4.01 |
| ADANF(ours) | 28.01 | 0.891 | 0.134 | 252.39 | 117.59 |

Table 3. Quantitative results of the proposed ADANF and the state-of-the-art methods for the LOLv2-synthetic datasets. ↑/↓ means that the larger/smaller the index value, the better/lower the quality. GFLOPs represents the Giga Floating Point Operations. Params represents the number of weight parameters.

| Method | LOLv2-Synthetic | | | Complexity | |
|-------------------|-----------------|-------|--------|------------|--------|
| | PSNR↑ | SSIM↑ | LPIPS↓ | FLOPs | Params |
| LIME [3] | 16.88 | 0.776 | 0.675 | - | - |
| RetinexNet [21] | 17.13 | 0.798 | 0.754 | 587.47 | 0.84 |
| DRBN [60] | 23.22 | 0.927 | - | 48.61 | 5.27 |
| KinD [43] | 16.26 | 0.591 | 0.435 | 34.99 | 8.02 |
| KinD++ [61] | - | - | - | - | 9.63 |
| EnlightenGAN [62] | 16.57 | 0.734 | - | 61.01 | 114.35 |
| MIRNet [63] | 21.94 | 0.846 | - | 785 | 31.76 |
| LLFlow [35] | 24.81 | 0.919 | 0.067 | - | 37.68 |
| LLFormer [16] | 28.01 | 0.927 | 0.061 | - | 24.55 |

Table 3. Cont.

| Method | LOLv2-Synthetic | | | Complexity | |
|----------------|-----------------|-----------------|--------------------|------------|--------|
| | PSNR \uparrow | SSIM \uparrow | LPIPS \downarrow | FLOPs | Params |
| Restormer [66] | 21.41 | 0.830 | - | 144.25 | 26.13 |
| SNR-Aware [31] | 27.79 | 0.941 | 0.054 | 26.35 | 4.01 |
| ADANF(ours) | 28.67 | 0.953 | 0.040 | 252.39 | 117.59 |

4.4. Comparisons with the State-of-the-Art Methods on Unpaired Datasets

In this section, we also conduct experiments on unpaired datasets. Due to the lack of reference images for comparison, we mainly used the NIQE to quantify the performance of each method and used the visual results for qualitative analysis. In terms of the NIQE indicator, the quantitative results for different datasets are shown in Table 4. The proposed ADANF shows better performance on the LIME, MEF, and VV datasets than other methods. On the DICM and NPE datasets, the proposed ADANF also has comparable performance.

Table 4. Quantitative results of the proposed ADANF and the state-of-the-art methods for unpaired DICM, LIME, MEF, NPE, and VV datasets. The evaluation index is NIQE.

| Methods | DICM | LIME | MEF | NPE | VV |
|-------------------|------|------|------|------|------|
| Zero-DCE [47] | 4.58 | 5.82 | 4.93 | 4.53 | 4.81 |
| EnlightenGAN [62] | 4.06 | 4.59 | 4.70 | 3.99 | 4.04 |
| RetinexNet [21] | 4.33 | 5.75 | 4.93 | 4.95 | 4.32 |
| KinD [43] | 3.95 | 4.42 | 4.45 | 3.92 | 3.72 |
| KinD++ [61] | 3.89 | 4.90 | 4.55 | 3.91 | 3.82 |
| DCC-Net [67] | 3.70 | 4.42 | 4.59 | 3.70 | 3.28 |
| ADANF(ours) | 3.90 | 3.78 | 3.59 | 4.24 | 3.14 |

4.5. Visualization

Visual results of different image enhancement methods on paired datasets. In order to verify that this method can generate better quality illumination-enhanced images, we compared some of the images generated by this method with the results generated by other low-light image enhancement algorithms. As shown in Figure 3, it can be seen that our ADANF can obtain a more realistic restoration effect. Compared with some methods, it has pictures with lower noise and more realistic colors. In addition, it has clearer details at the intersection of light and dark. These results can show that the module designed by this method is more complete and sufficient in extracting the features of the original image, promoting the final enhancement result to show more details in the transition area, thereby obtaining better enhancement results.

Visual results of different image enhancement methods on unpaired datasets. As can be seen from Figure 4, our method has better color performance in different scenarios. Compared with other methods, the image color obtained by this method is more realistic. It is neither too dark to see the details nor too bright to make the image color unrealistic. These visual results can show that the proposed method is effective not only in scenarios with paired datasets but also in complex scenarios with only unpaired datasets. Experimental results on multiple unpaired DICM, LIME, MEF, NPE, and VV datasets show that the proposed ADANF has good generalization ability.

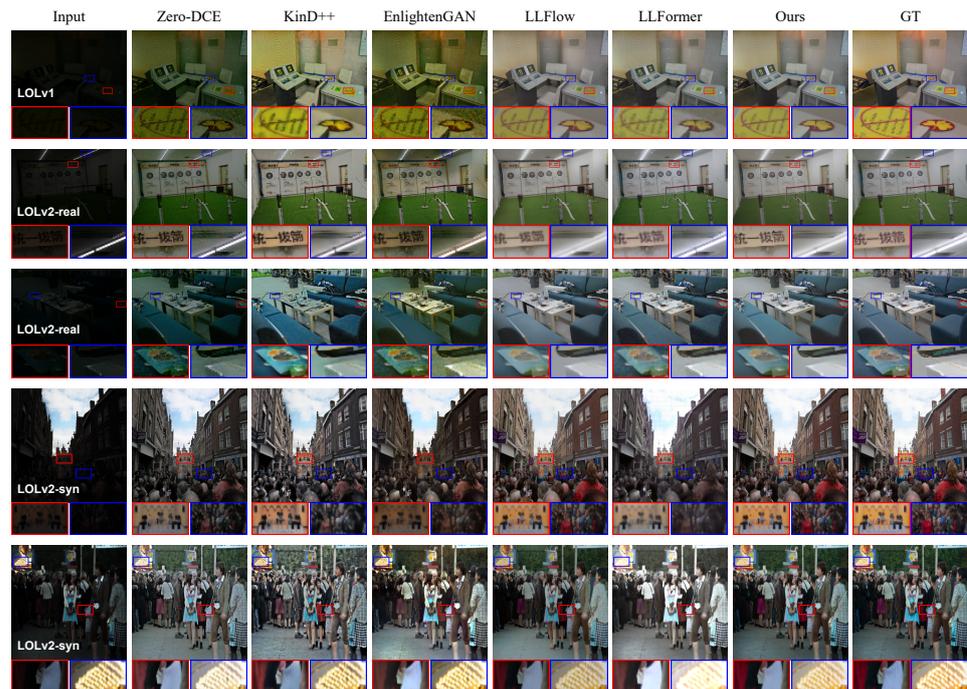


Figure 3. Some visualization results of the proposed ADANF and the recent state-of-the-art methods for the LOLv1, LOLv2-real, and LOLv2-synthetic datasets.

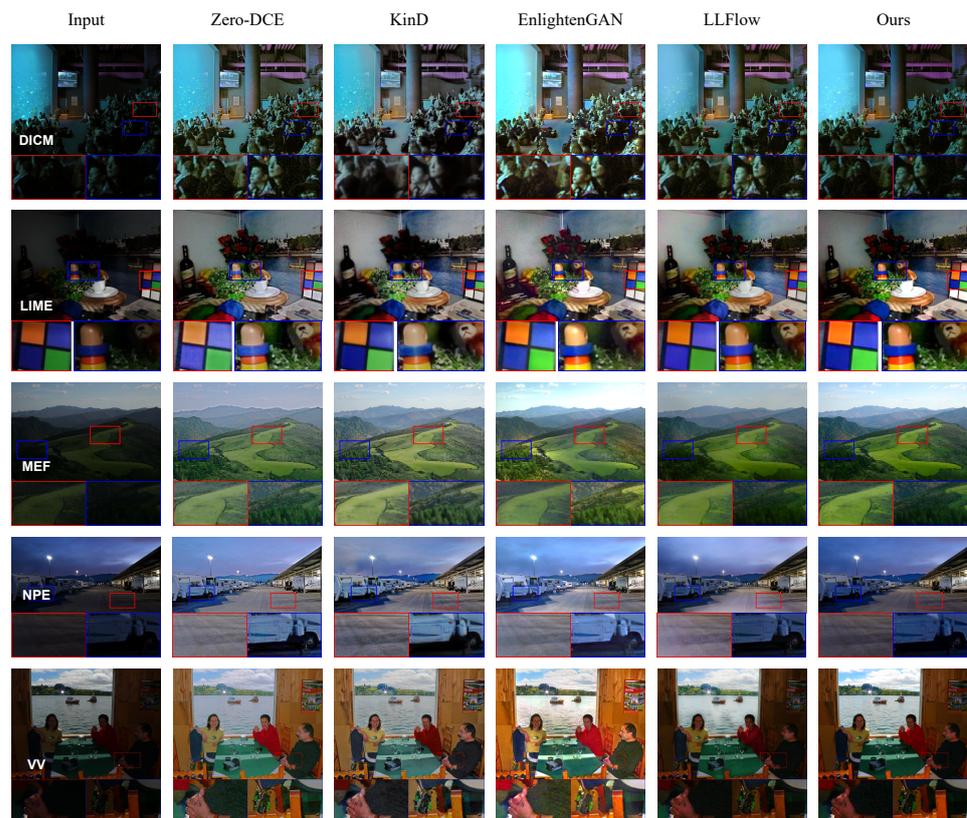


Figure 4. Some visualization results of the proposed ADANF and the recent state-of-the-art methods for the DICM, LIME, MEF, NPE, and VV datasets.

4.6. Ablation Study

In order to verify the effectiveness of the different modules designed by this method, we conducted ablation experiments on the LOLv1 dataset to test the effects of the introduced

adaptive dual aggregation encoder (ADAE) and gated multi-scale information transmitting module (GMITM). We replaced the ADAE in the proposed ADANF with multi-layer convolutions and removed the GMITM method as a baseline method. Then, the ADAE and GMITM modules are respectively added to the baseline method to conduct experiments. The experimental results are shown in Table 5. Compared with the baseline method, the introduction of the ADAE and GMITM alone can bring about improvements in the three evaluation indicators. The improvement brought by the introduction of the ADAE is that this module can fully utilize the potential features of the input image in both global and local aspects. The improvement obtained by further combining the GMITM and ADAE is because the image features extracted by ADAE are better transferred to the normalizing flow decoder to assist in image enhancement.

Table 5. Ablation studies of the proposed ADANF on the LOLv1 dataset. \uparrow/\downarrow means that the larger/smaller the index value, the better/lower the quality.

| ADAE | GMITM | PSNR \uparrow | SSIM \uparrow | LPIPS \downarrow |
|--------------|--------------|-----------------|-----------------|--------------------|
| | | 24.83 | 0.819 | 0.157 |
| \checkmark | | 26.05 | 0.822 | 0.134 |
| | \checkmark | 25.91 | 0.845 | 0.126 |
| \checkmark | \checkmark | 26.67 | 0.873 | 0.120 |

5. Conclusions

In this paper, we propose an adaptive dual aggregation network with normalizing flows for low-light image enhancement. First, an adaptive dual aggregation encoder is used to fully exploit the global properties and local details of the image to extract illumination-robust features. Next, after illumination-robust features are extracted, a reversible normalizing flow decoder is used to recover normally exposed images from these features. This step takes advantage of the inverse process capabilities of the normalizing flow decoder to reconstruct brighter, more detailed images from low-light images. Finally, a gated multi-scale information transmitting module is designed to introduce the multi-scale features of the adaptive dual aggregation encoder into the normalizing flow decoder. This step aims to further improve the quality of image enhancement by introducing multi-scale features. Extensive experiments on paired and unpaired datasets verify the effectiveness of the proposed ADANF. In the future, we will study lightweight low-light image enhancement networks to meet the needs of real-time low-light image processing applications.

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

Funding: This research was funded by the National Science Basic Research Plan in Shannxi Province of 379 China grant number 2023-JQ-QC-0714.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

References

- Zhang, K.; Yuan, C.; Li, J.; Gao, X.; Li, M. Multi-Branch and Progressive Network for Low-Light Image Enhancement. *IEEE Trans. Image Process.* **2023**, *32*, 2295–2308. [[CrossRef](#)]
- Fan, G.D.; Fan, B.; Gan, M.; Chen, G.Y.; Chen, C.L.P. Multiscale Low-Light Image Enhancement Network With Illumination Constraint. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 7403–7417. [[CrossRef](#)]

3. Guo, X.; Li, Y.; Ling, H. LIME: Low-Light Image Enhancement via Illumination Map Estimation. *IEEE Trans. Image Process.* **2017**, *26*, 982–993. [[CrossRef](#)]
4. Sun, H.; Li, Q.; Yu, J.; Zhou, D.; Chen, W.; Zheng, X.; Lu, X. Deep Feature Reconstruction Learning for Open-Set Classification of Remote-Sensing Imagery. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [[CrossRef](#)]
5. Li, C.; Guo, C.; Han, L.; Jiang, J.; Cheng, M.M.; Gu, J.; Loy, C.C. Low-Light Image and Video Enhancement Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 9396–9416. [[CrossRef](#)]
6. Brateanu, A.; Balmez, R.; Avram, A.; Orhei, C. LYT-Net: Lightweight YUV Transformer-based Network for Low-Light Image Enhancement. *arXiv* **2024**, arXiv:2401.15204.
7. Kong, N.S.P.; Ibrahim, H. Color image enhancement using brightness preserving dynamic histogram equalization. *IEEE Trans. Consum. Electron.* **2008**, *54*, 1962–1968. [[CrossRef](#)]
8. Ibrahim, H.; Pik Kong, N.S. Brightness Preserving Dynamic Histogram Equalization for Image Contrast Enhancement. *IEEE Trans. Consum. Electron.* **2007**, *53*, 1752–1758. [[CrossRef](#)]
9. Jeong, I.; Lee, C. An optimization-based approach to gamma correction parameter estimation for low-light image enhancement. *Multimed. Tools Appl.* **2021**, *80*, 18027–18042. [[CrossRef](#)]
10. Kim, W. Low-Light Image Enhancement: A Comparative Review and Prospects. *IEEE Access* **2022**, *10*, 84535–84557. [[CrossRef](#)]
11. Bertalmío, M.; Caselles, V.; Provenzi, E. Issues about retinex theory and contrast enhancement. *Int. J. Comput. Vis.* **2009**, *83*, 101–119. [[CrossRef](#)]
12. Ren, X.; Yang, W.; Cheng, W.H.; Liu, J. LR3M: Robust Low-Light Enhancement via Low-Rank Regularized Retinex Model. *IEEE Trans. Image Process.* **2020**, *29*, 5862–5876. [[CrossRef](#)]
13. Gu, Z.; Li, F.; Fang, F.; Zhang, G. A Novel Retinex-Based Fractional-Order Variational Model for Images With Severely Low Light. *IEEE Trans. Image Process.* **2020**, *29*, 3239–3253. [[CrossRef](#)] [[PubMed](#)]
14. Wang, Y.; Yu, Y.; Yang, W.; Guo, L.; Chau, L.P.; Kot, A.C.; Wen, B. Exposediffusion: Learning to expose for low-light image enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 12438–12448.
15. Rasheed, M.T.; Shi, D.; Khan, H. A comprehensive experiment-based review of low-light image enhancement methods and benchmarking low-light image quality assessment. *Signal Process.* **2023**, *204*, 108821. [[CrossRef](#)]
16. Wang, T.; Zhang, K.; Shen, T.; Luo, W.; Stenger, B.; Lu, T. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 2654–2662.
17. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [[CrossRef](#)]
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
19. Liu, X.; Ma, W.; Ma, X.; Wang, J. LAE-Net: A locally-adaptive embedding network for low-light image enhancement. *Pattern Recognit.* **2023**, *133*, 109039. [[CrossRef](#)]
20. 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](#)]
21. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep Retinex Decomposition for Low-Light Enhancement. In Proceedings of the British Machine Vision Conference 2018, Newcastle, UK, 3–6 September 2018.
22. Yang, W.; Wang, W.; Huang, H.; Wang, S.; Liu, J. Sparse Gradient Regularized Deep Retinex Network for Robust Low-Light Image Enhancement. *IEEE Trans. Image Process.* **2021**, *30*, 2072–2086. [[CrossRef](#)]
23. Guo, X.; Hu, Q. Low-light image enhancement via breaking down the darkness. *Int. J. Comput. Vis.* **2023**, *131*, 48–66. [[CrossRef](#)]
24. HU, X.; WANG, J.; XU, S. Lightweight and Fast Low-Light Image Enhancement Method Based on PoolFormer. *IEICE Trans. Inf. Syst.* **2024**, *107*, 157–160. [[CrossRef](#)]
25. 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.
26. Ma, L.; Liu, R.; Zhang, J.; Fan, X.; Luo, Z. Learning deep context-sensitive decomposition for low-light image enhancement. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 5666–5680. [[CrossRef](#)]
27. Liu, X.; Xie, Q.; Zhao, Q.; Wang, H.; Meng, D. Low-Light Image Enhancement by Retinex-Based Algorithm Unrolling and Adjustment. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, 1–14. [[CrossRef](#)]
28. 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, New Orleans, LA, USA, 18–24 June 2022; pp. 5901–5910.
29. Fan, M.; Wang, W.; Yang, W.; Liu, J. Integrating Semantic Segmentation and Retinex Model for Low-Light Image Enhancement. In Proceedings of the 28th ACM International Conference on Multimedia (MM 2020), Seattle, WA, USA, 12–16 October 2020; Association for Computing Machinery, Inc.: New York, NY, USA, 2020; pp. 2317–2325.
30. 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.

31. Xu, X.; Wang, R.; Fu, C.W.; Jia, J. SNR-aware 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. 17714–17724.
32. Li, J.; Feng, X.; Hua, Z. Low-light image enhancement via progressive-recursive network. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 4227–4240. [[CrossRef](#)]
33. Yang, S.; Zhou, D.; Cao, J.; Guo, Y. LightingNet: An Integrated Learning Method for Low-Light Image Enhancement. *IEEE Trans. Comput. Imaging* **2023**, *9*, 29–42. [[CrossRef](#)]
34. Zhang, Z.; Jiang, Z.; Liu, J.; Fan, X.; Liu, R. Waterflow: Heuristic normalizing flow for underwater image enhancement and beyond. In Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October–3 November 2023; pp. 7314–7323.
35. Wang, Y.; Wan, R.; Yang, W.; Li, H.; Chau, L.P.; Kot, A. Low-light image enhancement with normalizing flow. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual Event, 22 February–1 March 2022; Volume 36, pp. 2604–2612.
36. Lugmayr, A.; Danelljan, M.; Van Gool, L.; Timofte, R. Srfflow: Learning the super-resolution space with normalizing flow. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part V 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 715–732.
37. Jähne, B. *Digital Image Processing*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2005.
38. 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](#)]
39. Lee, C.; Lee, C.; Kim, C.S. Contrast Enhancement Based on Layered Difference Representation of 2D Histograms. *IEEE Trans. Image Process.* **2013**, *22*, 5372–5384. [[CrossRef](#)] [[PubMed](#)]
40. Gu, K.; Wang, S.; Zhai, G.; Ma, S.; Yang, X.; Lin, W.; Zhang, W.; Gao, W. Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure. *IEEE Trans. Multimed.* **2016**, *18*, 432–443. [[CrossRef](#)]
41. Kimmel, R.; Elad, M.; Shaked, D.; Keshet, R.; Sobel, I. A variational framework for retinex. *Int. J. Comput. Vis.* **2003**, *52*, 7–23. [[CrossRef](#)]
42. Liang, Z.; Liu, W.; Yao, R. Contrast Enhancement by Nonlinear Diffusion Filtering. *IEEE Trans. Image Process.* **2016**, *25*, 673–686. [[CrossRef](#)]
43. Zhang, Y.; Zhang, J.; Guo, X. Kindling the Darkness: A Practical Low-Light Image Enhancer. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 1632–1640.
44. Ren, W.; Liu, S.; Ma, L.; Xu, Q.; Xu, X.; Cao, X.; Du, J.; Yang, M.H. Low-Light Image Enhancement via a Deep Hybrid Network. *IEEE Trans. Image Process.* **2019**, *28*, 4364–4375. [[CrossRef](#)] [[PubMed](#)]
45. Xu, K.; Yang, X.; Yin, B.; Lau, R.W. Learning to restore low-light images via decomposition-and-enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 2281–2290.
46. Cai, J.; Gu, S.; Zhang, L. Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. *IEEE Trans. Image Process.* **2018**, *27*, 2049–2062. [[CrossRef](#)] [[PubMed](#)]
47. 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, 14–19 June 2020; pp. 1780–1789.
48. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
49. Yao, H.; Chen, R.; Chen, W.; Sun, H.; Xie, W.; Lu, X. Pseudolabel-Based Unreliable Sample Learning for Semi-Supervised Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–16. [[CrossRef](#)]
50. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
51. Kingma, D.P.; Dhariwal, P. Glow: Generative flow with invertible 1x1 convolutions. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 1–10.
52. Dinh, L.; Sohl-Dickstein, J.; Bengio, S. Density estimation using Real NVP. In Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.
53. He, Z.; Ran, W.; Liu, S.; Li, K.; Lu, J.; Xie, C.; Liu, Y.; Lu, H. Low-Light Image Enhancement with Multi-Scale Attention and Frequency-Domain Optimization. *IEEE Trans. Circuits Syst. Video Technol.* **2023**. [[CrossRef](#)]
54. Orhei, C.; Vasiu, R. An Analysis of Extended and Dilated Filters in Sharpening Algorithms. *IEEE Access* **2023**, *11*, 81449–81465. [[CrossRef](#)]
55. Lee, C.; Lee, C.; Kim, C.S. Contrast enhancement based on layered difference representation. In Proceedings of the 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; pp. 965–968.
56. Li, M.; Liu, J.; Yang, W.; Sun, X.; Guo, Z. Structure-Revealing Low-Light Image Enhancement Via Robust Retinex Model. *IEEE Trans. Image Process.* **2018**, *27*, 2828–2841. [[CrossRef](#)]
57. Wang, S.; Zheng, J.; Hu, H.M.; Li, B. Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. *IEEE Trans. Image Process.* **2013**, *22*, 3538–3548. [[CrossRef](#)]
58. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]

59. 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.
60. Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; Liu, J. From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
61. Zhang, Y.; Guo, X.; Ma, J.; Liu, W.; Zhang, J. Beyond Brightening Low-light Images. *Int. J. Comput. Vis.* **2021**, *129*, 1013–1037. [[CrossRef](#)]
62. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. *IEEE Trans. Image Process.* **2021**, *30*, 2340–2349. [[CrossRef](#)] [[PubMed](#)]
63. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Learning Enriched Features for Fast Image Restoration and Enhancement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 1934–1948. [[CrossRef](#)] [[PubMed](#)]
64. Yi, X.; Xu, H.; Zhang, H.; Tang, L.; Ma, J. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 12302–12311.
65. Yang, S.; Zhou, D.; Cao, J.; Guo, Y. Rethinking low-light enhancement via transformer-GAN. *IEEE Signal Process. Lett.* **2022**, *29*, 1082–1086. [[CrossRef](#)]
66. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5728–5739.
67. Zhang, Z.; Zheng, H.; Hong, R.; Xu, M.; Yan, S.; Wang, M. Deep Color Consistent Network for Low-Light Image Enhancement. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 1889–1898. [[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.