

Article

TSRNet: A Trans-Scale and Refined Low-Light Image Enhancement Network

Qi Mu *, Yueyue Ma, Xinyue Wang and Zhanli Li

College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710054, China; 21208088023@stu.xust.edu.cn (Y.M.); 21308223002@stu.xust.edu.cn (X.W.); lizl@xust.edu.cn (Z.L.)

* Correspondence: muqi@xust.edu.cn

Abstract: Retinex-based deep learning methods show good low-light enhancement performance and are mainstream approaches in this field. However, the current methods for enhancing low-light images are insufficient in accurately separating illumination and comprehensively restoring degraded information, especially in images with uneven or extremely low illumination levels. This situation often leads to the over-enhancement of bright regions, a loss of detail, and color distortion in the final images. To address these issues, we improved three subnetworks in the classic KinD network, and proposed a trans-scale and refined low-light image enhancement network. Compared with KinD, our method shows more precise image decomposition performance, enhancing the expressiveness of the reflection and illumination components in order to better depict image details, colors, and lighting information. For reflectance restoration, we use a U-shaped network for cross-scale denoising, incorporating attention mechanisms and a color saturation loss to restore image textures and colors. For light adjustment, we apply fine-grained light adjustment approaches to simultaneously enhance brightness in dark areas and prevent excessive enhancement in bright areas. The experimental results demonstrate that with the LOL dataset, the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) of TSRNet are improved by 2–31% and 5–34%, respectively, when compared with the mainstream methods.

Keywords: nonuniform low illumination; extremely low illumination; trans-scale; refined; U-shaped network



Citation: Mu, Q.; Ma, Y.; Wang, X.; Li, Z. TSRNet: A Trans-Scale and Refined Low-Light Image Enhancement Network. *Electronics* **2024**, *13*, 950. <https://doi.org/10.3390/electronics13050950>

Academic Editor: Guido Masera

Received: 1 February 2024

Revised: 20 February 2024

Accepted: 26 February 2024

Published: 29 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

Although images inherently contain rich information, images are often acquired under uneven or extremely low illumination in practical scenarios, such as during nighttime surveillance, in backlit environments, or in poorly lit rooms. Consequently, these images may exhibit low brightness, low contrast levels, significant noise, and blurred details [1], not only resulting in a suboptimal visual experience for viewers, but also impeding the performance on advanced visual tasks, such as facial detection and target recognition, in low-light environments [2]. Therefore, to make the buried details and color visible, to enhance human visual perception, and to accelerate advancement in the execution of advanced visual tasks, low-light image enhancement is demanded.

Traditional low-light image enhancement methods can be divided into two categories as follows: mapping-based methods [3,4] and retinex-based methods [5–7]. Mapping-based methods [8,9] are designed to improve image contrast, but, although they can effectively enhance image brightness, the resulting images may contain obvious noise and blurred details. Retinex-based methods [5,6] focus on eliminating the influence of illumination, and can therefore reduce noise to some extent when compared to mapping-based methods. However, the current illumination estimation methods are not sufficiently accurate, resulting in a loss of detail and color distortion in the final results. Figure 1a,b show original images acquired under uneven and extremely low illumination, respectively.

As shown in Figure 1c–f, representative mapping-based and retinex-based algorithms, histogram equalization (HE) [10], and multiscale retinex (MSR) [11], respectively, produced poor enhancement effects on Figure 1a,b.

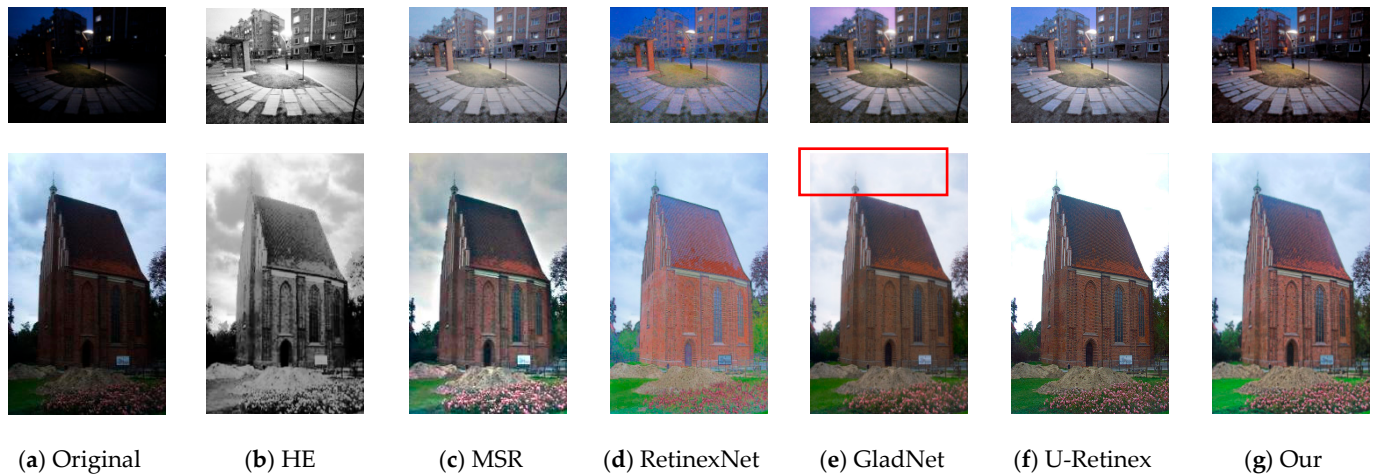


Figure 1. The enhanced results of different algorithms in images with uneven or extremely low illumination; (g) are our enhanced results.

Compared with traditional methods, methods based on deep learning have been demonstrated to achieve better enhancement effects. For instance, Figure 1d,e show the enhancement results of two representative deep learning-based networks, RetinexNet [12] and GladNet [13]. The enhancement results are evidently brighter and free of severe halos (as observed around the house in Figure 1c) compared with those produced by the traditional methods. However, due to the lack of a trans-scale and refined network structure, these methods still face the three following challenges when processing images captured under uneven or extremely low illumination: detail blurring, color distortion, and the over-enhancement of bright regions. For example, regions with different levels of exposure in an image will contain varying levels of noise, with the noise in dark regions being much more severe than that in bright regions. Consequently, RetinexNet’s uniform denoising strategy cannot effectively eliminate noise in images with uneven illumination. Moreover, due to its lack of color constraints, RetinexNet suffers from noticeable noise and color distortion in its enhancement results, as shown in Figure 1d. Similarly, GladNet’s global illumination estimation method cannot finely adjust the brightness levels of images with uneven illumination, which can lead to over-enhancement in bright regions, as shown in the region marked with a red box in Figure 1e.

To address the aforementioned challenges, we propose a low-light enhancement network based on the concepts of trans-scale processing and refinement, which can maintain details, restore color, and adaptively adjust unevenly illuminated regions. Figure 1g shows the results of our method. Noise is effectively eliminated, while clear details are preserved. Moreover, the color appears realistic and bright. The over-enhancement of bright regions does not occur, and no artifacts appear at the boundaries between regions with different illumination levels. In comparison to state-of-the-art Retinex-based methods (as shown in Figure 1f), TSRNet also shows satisfactory color restoration and illumination adjustment results.

The main contributions of this paper are as follows:

- (1) We added all cross-level connections based on U-shaped networks for image decomposition.
- (2) In the reflectance refinement restoration network, in addition to the use of the U-shaped network for cross-scale denoising, we incorporate attention mechanisms and a color saturation loss to obtain clear details and natural colors.

- (3) In the lighting adjustment network, we train a detailed factor that can adaptively adjust the brightness of each pixel. Importantly, the brightness adjustment range of this factor is not limited.

Compared with the classic KinD approach, in terms of image decomposition, TSRNet enhances the expressiveness of the reflection and illumination to better depict image details, colors, and lighting information, thus providing a good foundation for the subsequent image detail restoration process. In terms of reflectance restoration, TSRNet balances noise reduction with detail preservation, while also obtaining natural and visually appealing image colors. In terms of illumination adjustment, our factor is more fine-grained than previous approaches, thus enhancing the generalizability of the illumination adjustment function.

2. Related Work

In recent years, researchers have proposed a series of schemes for enhancing low-light images. In the following section, we will provide a brief overview of classical algorithms that are closely related to the topic of this paper.

2.1. Retinex-Based Traditional Methods

Retinex theory [14] posits that an image is a product of the illumination and reflectance, as shown in Formula (1), where (x, y) denotes the two-dimensional coordinates of an image pixel, and $L(x, y)$, $R(x, y)$, and $I(x, y)$, respectively, represent the original image, reflectance, and illumination. Retinex-based image enhancement methods [6,7] estimate the illumination through some prior or regularization and remove it from the original image in order to eliminate the adverse effects of illumination on the image. However, it is difficult to accurately estimate the illumination. The illumination estimated by the existing prior and regularization is not accurate [15]. Thus, the influence of illumination in the resulting image is not completely eliminated, which results in detail loss and color deviation. Therefore, finding an effective prior or regularization is a key problem that still needs to be explored for such methods [15].

$$L(x, y) = R(x, y) \bullet I(x, y) \quad (1)$$

2.2. Deep Learning-Based Methods

In recent years, deep learning-based methods for low-light image enhancement have demonstrated significant advancements and outstanding achievements. In 2017, Lore et al. [16] introduced LLNet, which was the pioneering work to apply deep learning to image enhancement, and yielded remarkable results. LLNet innovatively uses deep sparse denoising auto-encoders to construct a network for both the contrast enhancement and denoising, whereas the details were not taken into account, resulting in blurred details. For display-rich details, the GladNet, proposed by Wang et al. [13] in 2018, designs a network named detail reconstruction, which has achieved remarkable success in retaining image details. In addition, the global illumination estimation method used by GladNet can effectively enhance the brightness of low-light images, whereas, when processing the uneven illumination images, GladNet applies the same brightness enhancement to different brightness regions, which inevitably leads to over-enhancement in bright regions and a loss of details.

2.3. Retinex-Based Deep Learning Methods

Unlike the direct adjustment of the contrast and brightness of images based on subjective visual perception in Section 2.2, several scholars argue that applying the Retinex theory to image enhancement has a more robust theoretical foundation. Therefore, scholars have tried to combine Retinex with CNN (convolutional neural network) to obtain better illumination and reflectance with the powerful learning ability of CNN. The enhancement steps for this type of method can be summarized as follows: Using a pair of low/normal-light

images as the network input, a pair of decomposed reflectance and illumination images are obtained. Then, enhancements are separately applied to the reflectance and illumination components, and, finally, the fusion of the enhanced illumination and reflectance produces the final enhanced image.

In 2018, Wei et al. [12] proposed RetinexNet, based on the Retinex theory, which decomposes the input image into reflectance and illumination via a decomposition network. Subsequently, it executes the brightness enhancement of illumination and the denoising of reflectance. Based on this, this method can excellently improve the brightness in extremely low illumination images, and will not lead to over-enhancement in unevenly illuminated images. However, the classic denoising method, named BM3D (block-matching and 3D filtering) [17], used in RetinexNet is not entirely suitable for low-light images. This is because the noise in dark regions is more severe than in bright regions, and the uniform denoising proposed in BM3D is unable to completely eliminate noise in low-light images. Thus, RetinexNet does not have great denoising abilities. In addition, it also lacks color constraints, making the color extremely unnatural in the resulting image.

Inspired by RetinexNet, Zhang et al. [18] proposed KinD in 2019. They designed three independent subnetworks for image decomposition, reflectance restoration, and illumination adjustment, based on the divide and conquer principle. Compared with RetinexNet, the reflectance restoration network in KinD can more effectively eliminate noise. Furthermore, the illumination adjustment network in KinD incorporates an adjustment factor that can adjust the brightness to a certain degree. However, KinD has the following three shortcomings: (1) Shallow-level information is often obtained early in the network. KinD uses U-Net for image decomposition. During the layer-by-layer convolutions, shallow-level information may be lost [15], resulting in the loss of detail in the decomposed reflectance image, and ultimately leading to a blurry enhanced image. (2) KinD utilizes U-Net for denoising in the reflectance restoration network. However, during this process, KinD does not account for the detail loss due to denoising. (3) KinD roughly adjusts the illumination levels with the illumination adjustment network; that is, $\text{ratio} > 1$ and $\text{ratio} \leq 1$ [18]. As a result, the enhancement outcomes are prone to issues such as over-enhancement or insufficient improvement in the overall image brightness.

2.4. U-Shaped Network

The Retinex-based deep learning methods often use convolution to extract the reflectance. The shallow convolution layer can extract basic edge and texture features. As the network deepens, it combines shallow features into basic shapes and local features, eventually forming a complete representation of the entire object. However, during this process, there is a risk of losing shallow texture details, which leads to unclear reflectance and subsequently affects the recovery of reflectance.

The U-shaped network can integrate low-level and high-level information, which is beneficial to low-light image enhancement. Consequently, more and more researchers apply U-Net [19] or U-shaped networks to image decomposition or denoising [15]. In 2021, Lv and Sun et al. [20] utilized a U-shaped network to design a decomposition network for obtaining illumination and reflectance. Based on this, they achieved satisfactory decomposition results. Additionally, after enhancing the brightness of the image, the network also utilized U-Net for detail restoration. In 2022, Zhao et al. [21] proposed a unified depth framework for low-light image enhancement. It uses ResNet [22] and U-Net for image decomposition, respectively, and experiments showed that U-Net was more capable of obtaining a satisfactory decomposition result.

2.5. Obtain Inspiration

In summary, the above algorithm can effectively process images captured under uneven or extremely low illumination conditions. However, during denoising, it can easily cause the blurring of image edges, as exemplified by LLNet, RetinexNet, and KinD. Moreover, when enhancing the brightness of images with uneven illumination levels, these

methods roughly adjust the illumination level using one or two brightness adjustment ranges, as exemplified by GladNet and KinD. There is a risk of under-enhancing illumination levels in dark regions or over-enhancing illumination levels in bright regions, which can lead to a loss of detail. In addition, these algorithms still have shortcomings concerning color restoration, as the enhanced results exhibit color distortion. In previous research, the effectiveness of the Retinex theory for low-light enhancement has been demonstrated in both theory and practice, and the inherent structure of the U-shaped network is beneficial for the decomposition and enhancement of low-light images. Based on this, we improve the performance of Retinex-based deep learning methods as follows: For image decomposition, we improved the performance of the U-shaped network to achieve more refined decomposition results. During denoising, we apply a U-shaped network, incorporate an attention mechanism [23] to preserve details, and use a color saturation loss function to prevent color distortion. To adjust the lighting levels, we designed a detailed factor to adjust the illumination at the pixel level, and the brightness adjustment range of this factor was not limited.

3. Materials and Methods

Based on the survey results, we applied the classic KinD network framework, as shown in Figure 2. We improved the three subnetworks for image decomposition, reflectance component recovery, and lighting adjustment, proposing a deep neural network called TSRNet.

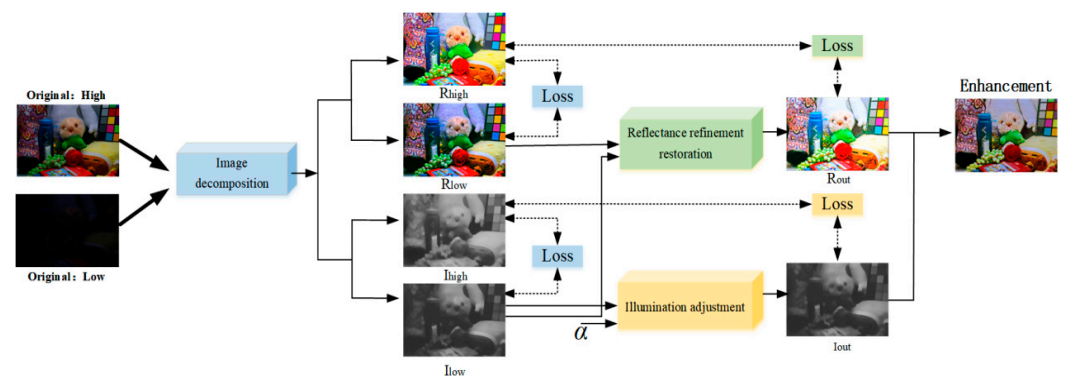


Figure 2. Network framework diagram.

3.1. Image Decomposition Network

The architecture of the image decomposition network is shown in Figure 3. Through decomposition, the more information contained in the reflectance, the better. Thus, we add all cross-level connections based on U-shaped nets to extract the reflectance with refined precision. Through decomposition, the illumination should reflect changes in light and maintain consistency with reflectance in terms of structure. So, we designed a network branch to extract the illumination. To ensure the accuracy of the decomposition results, we use composite loss function to constrain the decomposition network.

During the layer-by-layer convolutions, there is a risk of losing shallow-level information, resulting in the loss of detail in the decomposed reflectance, and ultimately leading to a blurry enhanced image. Research has found that U-Net++ [24], commonly used for image segmentation, cleverly adds skip connections between non-adjacent convolutional layers, achieving a win-win situation in extracting both detailed features and global features. Therefore, we use the same design to fuse features from different levels of depth and shallowness. It not only extracts rich abstract information in high-level, but also avoids the loss of textures in the shallow level, laying the foundation for the recovery of the reflectance. In addition, we also used the reflectance similarity loss in the decomposition

network, aiming to make the reflectance of low-light images as close as possible to that of normal-light images.

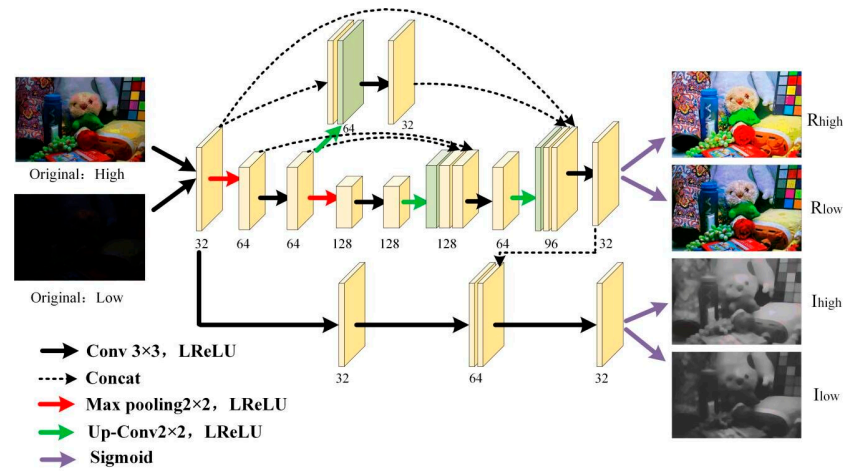


Figure 3. Image decomposition network structure.

According to Retinex [12], the illumination should be smooth within continuous regions, which means that light changes should only occur at the edges of objects. Therefore, when extracting illumination, we perform three 3×3 convolutions with the LReLU for feature extraction. Specifically, the input to the third convolution layer consists of two feature maps. The first one is the output of the second convolution layer, and the second one is the final feature map from the reflectance extraction branch. This is completed to allow the reflectance with clear edges, to help smooth out texture in the illumination, and to maintain the consistency in the structure of both components. Simultaneously, we incorporate a smoothness loss to regulate the illumination during this process, with the objective of achieving improved smoothness in the non-edge regions of the image.

The composite loss function L_{dec} consists of the four following terms: reconstruction loss L_{res} , reflectance similarity loss L_{rs} , illumination smoothing loss L_{is} , and mutual consistency loss L_{mc} .

$$L_{dec} = L_{res} + \alpha_{rs}L_{rs} + \alpha_{is}L_{is} + \alpha_{mc}L_{mc} \quad (2)$$

where α_{rs} , α_{is} , and α_{mc} denote the coefficients to balance the L_{rs} , L_{is} , and L_{mc} losses, whose values are taken as 0.01, 0.15, and 0.2, respectively, when training the network model [18].

L_{res} is used to ensure the accuracy of the decomposition. The reconstructed image of the decomposed reflection and illumination components should differ minimally from the original input image as follows:

$$L_{res} = \|S_{low} - R_{low} \bullet I_{low}\|_1 + \|S_{high} - R_{high} \bullet I_{high}\|_1 \quad (3)$$

where S_{low} and S_{high} represent the input normal and low light image. R_{low} and I_{low} represent the reflection and illumination decomposed from normal-light images. R_{high} and I_{high} represent the reflection and illumination decomposed from low-light images. And $\|\bullet\|_1$ means the $\uparrow_1$ norm.

The reflection component is unaffected by changes in illumination. Therefore, L_{rs} is used to ensure that the reflection component extracted from low-light images is as consistent as possible with that extracted from normal-light images.

$$L_{rs} = \|R_{high} - R_{low}\|_1 \quad (4)$$

In illumination smoothing loss, the input image is employed to guide the generation of the illumination component. In the regions of the edges in the input image, the illumination should vary, while in the flat regions, the illumination should be smooth. The structure

of L_{is} is depicted in Equation (5). L_{is} continuously decreases during the training process. Therefore, when ∇S is large, indicating image edges, the constraint on ∇I becomes smaller, preserving image information; conversely, when ∇S is small, indicating smooth regions, the constraint on ∇I increases, smoothing image information.

$$L_{is} = \left\| \frac{\nabla I_{low}}{\max(|\nabla S_{low}|, \rho)} \right\|_1 + \left\| \frac{\nabla I_{high}}{\max(|\nabla S_{norm}|, \rho)} \right\|_1 \quad (5)$$

where ∇ stands for the first order derivative operator, containing ∇_x (horizontal) and ∇_y (vertical) directions. ρ is a non-zero constant ensuring the avoidance of the zero denominator, set to 0.01. And $|\bullet|$ means the absolute value operator.

L_{mc} aims to make the structures of I_{low} and I_{high} as consistent as possible.

$$L_{mc} = \|\kappa \bullet \exp(-c \bullet \kappa)\|_1 \quad (6)$$

$$\kappa = |\nabla I_{low}| + |\nabla I_{high}| \quad (7)$$

where c is a parameter used to control the shape of the function, with a value of 10 during model training. And κ represents the sum of the gradients of the illumination between the high and low light images [18].

3.2. Reflectance Refinement Restoration Network

The reflectance, obtained from the image decomposition network, contains information such as noise, texture, and color. And the presence of noise may interfere with details and color. Thus, denoising is very important. However, while denoising, it is easy to blur the details, so the primary goal of our restoration network is to effectively denoise while retaining details. Additionally, natural and bright colors are crucial for achieving an optimal visual experience. Therefore, restoring image color is the second goal of our restoration network. Based on this, we design a reflectance refinement restoration network. The architecture of the restoration network is shown in Figure 4.

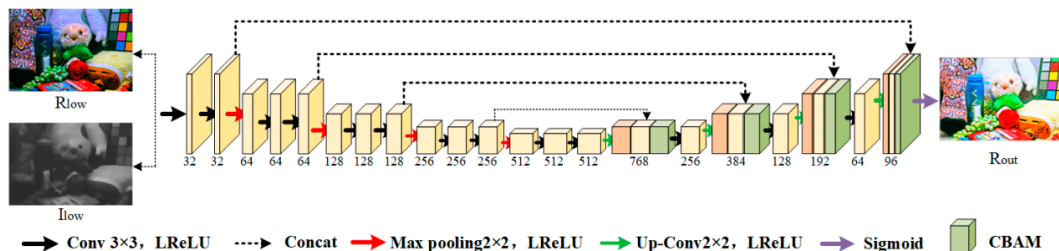


Figure 4. Reflectance refinement restoration network structure.

Denoising and retaining details. Upon analyzing the decomposed reflectance, some regions exhibit mild degradation, with isolated noise points as shown in the red box in Figure 5b, while other regions suffer from severe degradation, with noise distributed in block patterns as illustrated in the yellow box on the right in Figure 5d. For point noise, due to the high correlation among neighboring image pixels, restoration can be achieved through convolution operations utilizing the information from surrounding pixels. We choose the highly computationally efficient 3×3 convolution [25] for feature extraction. For block noise, expanding the convolution kernel size is a common way to achieve a large receptive field. However, it is easy to lead to a significant increase in the number of parameters and computational costs, which is not a preferable option for us. Therefore, we choose to use down-sampling to obtain a larger receptive field, and still use 3×3 convolutions to trans-scale eliminate the block noise. Integrating the CBAM in the up-sampling and concatenating the feature map between down-sampling and up-sampling is also aimed at enhancing details.

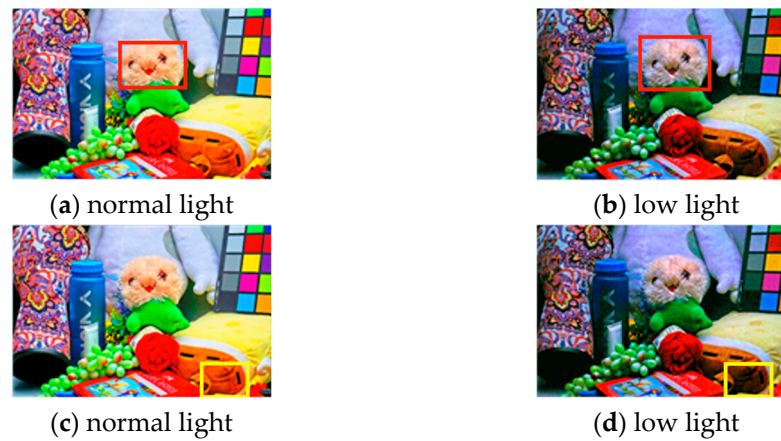


Figure 5. Noise distribution characteristics. (a,c) are images without degradation. They are the reference of (b,d).

Color recovery. A color saturation loss function [26] is employed to enforce color constraints. This loss function not only assesses the dissimilarities in color between paired reflectance samples, but also concurrently adjusts the brightness and contrast to ensure accurate color representation.

We defined a composite loss function L_{re} which consists of the four following terms: mean squared error loss L_{square} , gradient domain loss L_{grad} , structural similarity loss L_{ssim} , and color saturation loss L_{rgb} . R_{out} represents the recovered reflectance. R_{high} is the reference image. These can make R_{out} approach R_{high} in terms of edge details, brightness, contrast, and structure as follows:

$$L_{re} = L_{square} - L_{ssim} + L_{grad} + \lambda_{rgb} L_{rgb} \quad (8)$$

$$L_{square} = \left\| R_{out} - R_{high} \right\|_2^2 \quad (9)$$

$$L_{grad} = \left\| \nabla R_{out} - \nabla R_{high} \right\|_2^2 \quad (10)$$

$$L_{ssim} = SSIM(R_{out}, R_{high}) \quad (11)$$

where $\left\| \cdot \right\|_2^2$ means the l_2 norm, indicating the absolute error between the recovered reflectance and R_{high} (reference image). If the error increases, the mean squared loss increases quadratically. λ_{rgb} is set to 0.5. The value of λ_{rgb} is determined through experimentation. We conducted experiments by setting λ_{rgb} to 1, 0.8, 0.5, 0.3, and 0.1, respectively, all under the same conditions. In the last batch of the last epoch under different values, we randomly sampled 10 instances to compare their respective network losses (as shown in Figure 6a) and the time required to complete the restoration network training (as shown in Figure 6b). When λ_{rgb} is 0.5 (red curve), the average loss is the lowest and the required time is the least, therefore achieving the best model performance.

L_{rgb} can not only evaluate the color difference between the low-light reflection and the reference image, but also correct the brightness and contrast simultaneously. The main formulas of L_{rgb} are as follows:

$$L_{rgb}(R_{out}, R_{high}) = \left\| R'_{out} - R'_{high} \right\|_2^2 \quad (12)$$

$$R'(i, j) = \sum_{k, l} R(i + k, j + l) \cdot G(k, l) \quad (13)$$

$$G(k, l) = A \exp\left(-\frac{(k - \mu_x)^2}{2\delta_x} - \frac{(l - \mu_y)^2}{2\delta_y}\right) \quad (14)$$

where R'_{out} and R'_{high} are the blurred images of R_{out} and R_{high} , respectively, after Gaussian filtering $G(k, l)$. The specific parameters are set to $A = 0.053$, $\mu_{x,y} = 0$, $\delta_{x,y} = 0$ [26].

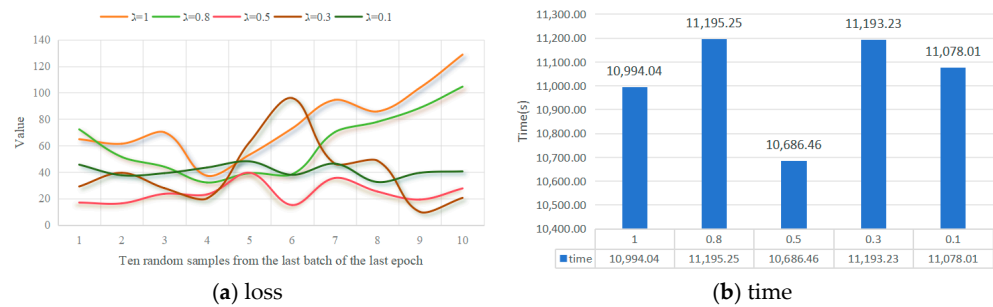


Figure 6. The impact of different values of λ_{rgb} on the restoration network loss value and time.

3.3. Illumination Adjustment Network

To accommodate the diverse requirements for various brightness levels in terms of light adjustment, we introduce an illumination adjustment ratio map α , as shown in Formula (15). The adjustment network is trained in a data-driven manner, with the illumination of normally lit images serving as the learning target. To guide this training process, a loss function is employed to impose constraints and optimize the performance of the network. By calculating the adjustment ratios between low-light and normal-light conditions on a per-pixel basis, the network can adaptively adjust to various brightness values, demonstrating strong generalization performance. Therefore, our network can enhance the brightness in dark regions, while avoiding over-enhancement in bright regions when processing images with uneven illumination. In addition, we apply the idea of a relative position to concatenate the feature maps before and after convolution to participate together in the next convolutional layer, thereby enabling the feature information to be fully utilized.

$$\alpha = \frac{I_{high}}{I_{low}} \quad (15)$$

The architecture of the illumination adjustment network is shown in Figure 7, which consists of five identical convolutional layers and one sigmoid layer. Combined with the idea of a relative position [22], we concatenate the feature maps before and after convolution to participate in the next convolutional layer together, fully utilizing the feature information.

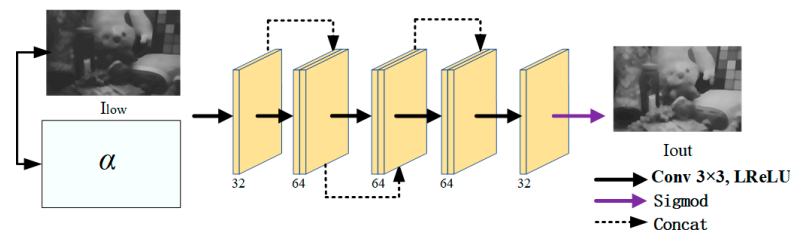


Figure 7. Illumination adjustment network structure.

The loss L_{adj} of this network consists of the mean squared error loss and gradient loss, where I_{out} represents the adjusted illumination,

$$L_{adj} = L'_{square} + L'_{grad} \quad (16)$$

$$L'_{square} = \left\| I_{out} - I_{high} \right\|_2^2 \quad (17)$$

$$L'_{grad} = \left\| \nabla I_{out} - \nabla I_{high} \right\|_2^2 \quad (18)$$

where $\| \cdot \|_2^2$ represents the absolute error between the recovered illumination and I_{high} (reference image). If the error increases, the mean squared loss increases quadratically.

3.4. Ablation Experiment

To verify the necessity of the operations used to address the over-enhancement of bright regions, a loss of details, and distorted color in the final image, we will conduct ablation experiments, including the U-Net++ network in the image decomposition network, color loss L_{rgb} and the CBAM attention mechanism in the reflection component network, and the light adjustment factor in the light adjustment network. In particular, because some algorithms commonly use U-Net for image decomposition [19,20], we compared the enhancement effects of using U-Net++ and U-Net separately in the decomposition network. The design of the ablation study is shown in Table 1, and the final enhanced results are shown in Figure 8.

Table 1. Ablation study design.

	U-Net++ for Decomposition	U-Net for Decomposition	CBAM	L_{rgb}	α
TSRNet	✓	-	✓	✓	✓
with U-Net	×	✓	✓	✓	✓
without CBAM	✓	-	×	✓	✓
without L_{rgb}	✓	-	✓	×	✓
without α	✓	-	✓	✓	×

Note: The last three rows are to verify the effectiveness of CBAM, L_{rgb} , and α , using subtraction on the basis of TSRNet and represented by “×” in the table. “✓” means use it.

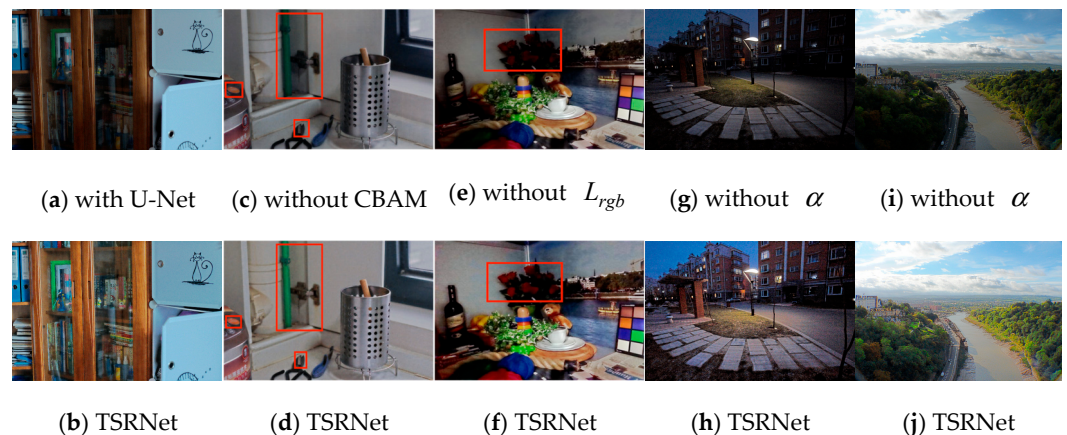


Figure 8. The enhancement results of the ablation experiment.

Figure 8 illustrates the enhancement results obtained before and after utilizing key points. Evidently, in Figure 8a,b, TSRNet, using U-Net++ for decomposition, exhibits superior enhancement effects over U-Net, characterized by enhanced levels of detail, heightened color vibrancy, and a notable preservation of fine details. Figure 8c,d show the experimental results before and after using CBAM. From the red box area, it can be observed that the image obtained by TSRNet with CBAM has clearer details. Figure 8e,f display the experimental results before and after using L_{rgb} . From the red box area, it can be seen that the image obtained by TSRNet with L_{rgb} has brighter colors. These improvements are particularly evident within the red roses and the white wall. Similarly, in Figure 8g–j, the results obtained after applying α exhibit outstanding performances in both bright

and dark regions. Table 2 shows the quantitative comparison of results in Figure 8. We employed a non-reference image quality assessment method (NIQE) [27], commonly used in the field of image enhancement at this stage, to evaluate the image quality. The smaller the NIQE value, the higher the quality of the image. From Table 2, it can be seen that the NIQE of enhancement results from TSRNet is always better than the images against which it is compared.

Table 2. The quantitative comparison of the results of the ablation experiments using the NIQE↓ index. Red is the optimal value.

Metrics	(a)	(c)	(e)	(g)	(i)
NIQE↓	5.0439	6.4148	3.5612	2.6991	2.6003
Metrics	(b)	(d)	(f)	(h)	(j)
NIQE↓	3.9901	5.5811	3.1583	2.5693	2.0392

4. Analysis of Experimental Results

4.1. Experimental Setup and Model Training

We use the LOL dataset [12] as our training dataset, which includes 500 pairs of low/normal light images. We use 485 images as the training set and the remaining 15 images as the testing set. The hyperparameters for the different subnetworks during training are shown in Table 3.

Table 3. The hyperparameter settings of the network model.

Network	Epoch	Batch_Size	Patch_Size	Optimizer	Parameter
1	2000	10	48 × 48	Adam	LR: 0.0001
2	1000	4	384 × 384		betal: 0.9 0.999
3	2000	10	48 × 48		epsilon: 1×10^{-8}

Note: “1” is Image Decomposition. “2” is Reflectance Refinement Restoration. “3” is Illumination Adjustment.

We evaluated our method on one reference dataset (LOL) and four non-reference datasets (VV [28], MEF [29], DICM [30], and LLIV-Phone-imgT [15]). For the evaluation, we selected a reference image quality assessment, namely PSNR [31] and SSIM [32], which measure image quality in terms of the signal-to-noise ratio, brightness, contrast, and structure. Additionally, we employed a non-reference image quality assessment method (NIQE) [27], commonly used in the field of image enhancement at this stage, to evaluate the image quality. The state-of-the-art methods of RetinexNet [12], GladNet [13], KinD [18], Zero-DCE [33], LLFlow [34], DRLIE [35], and URetinex [36] are involved as the competitors. All algorithms were run in the same environment, as shown in Table 4.

Table 4. The experimental environment of the TSRNet.

Equipment	Model
CPU	Intel(R) Core(TM) i7-8565 CPU @1.80GHz
GPU	NVIDIA Tesla P100
Operating system	Windows 10, 64bit
Experiment platform	Colaboratory (Colab)

4.2. Subjective Evaluation

Figure 9 displays the enhancement results of various algorithms for the ‘Home’ image from the testing set. Figure 10 shows the zoomed-in details of the regions marked with red boxes in Figure 9. RetinexNet yields a significant improvement in image brightness and preserves rich detailed information. However, the overall image exhibits obvious grain noise (as exemplified by the purple blanket in Figure 10c). GladNet has a better denoising

ability than RetinexNet. In the enhanced results produced by GladNet, the noise particles are reduced, but not completely eliminated. KinD effectively eliminates particle noise, but the texture of the image becomes blurry. The images enhanced by LLFlow and URetinex are very clear (as shown in Figure 9g,i). But, from comparisons with Figures 9b and 10b, it can be seen that the enhanced images produced by GladNet, KinD, LLFlow, DRLIE, and URetinex exhibit color deviations. In particular, the purple blanket becomes gray (as shown in Figure 10d,e,g–i). Zero-DCE fails to effectively enhance the brightness, and the enhanced result is blurry. From these comparisons, it is evident that our TSRNet exhibits remarkable efficacy in enhancing image brightness while effectively eliminating noise. Our enhanced image shows clear details and correct colors, which are very close to those of the reference image.

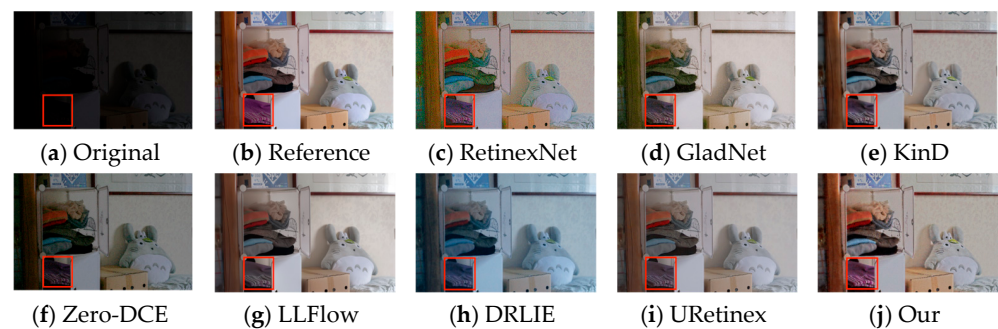


Figure 9. Comparison results of different algorithms in the Home image.

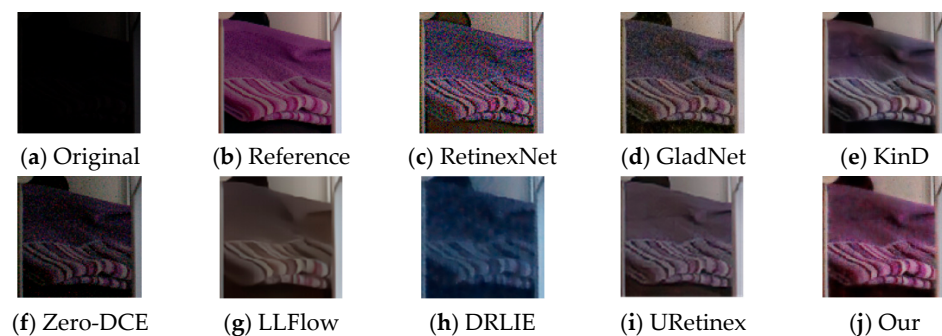


Figure 10. The zoomed-in details in the regions marked with red boxes in Figure 9.

Figure 11 presents the enhancement results of several images sourced from the VV, DICM, MEF, and LLIV-Phone-imgT datasets. From the results, the enhancement results from RetinexNet, GladNet, KinD, and Zero-DC, have the problems of noticeable noise, color deviation, excessive enhancement, and a loss of detail either individually or simultaneously. LLFlow and DRLIE exhibit the best enhancement results on natural landscape images (as shown in House). However, they have shown severe over-enhancement when processing other images. For example, the enhanced images of the girl and the woman, produced by DRLIE, and the tree in the night, produced by LLFlow, show excessive enhancement. Furthermore, the enhanced images of the girl, the woman, and the factory enhanced by DRLIE also exhibit incorrect color. When enhancing images with extremely low illumination, the results from URetinex are very clear. However, when enhancing images with uneven low illumination, it exhibits a significant issue of over-enhancement in bright regions. From these comparisons, in the image of the girl, an image with uneven illumination, our approach successfully enhances the brightness of backlit regions, while avoiding the over-enhancement of bright regions and preserving clear details (as seen in the girl's face, hair, and hand, which are very clear). In the image of the woman, TSRNet exhibits the most aesthetically pleasing effect for the sunset. In LLIV-Phone-imgT, compared to other

methods, we have clearer details and brighter and more natural color. Thus, the overall effect of TSRNet is the best.



Figure 11. Experimental comparison of different algorithms on VV, DICM, MEF, and LLIV-Phone-imgT.

4.3. Objective Evaluation

Table 5 shows the average of the PSNR, SSIM, and NIQE values of different methods, based on the testing set. Table 6 shows the NIQE of images from VV, DICM, MEF, and LLIV-Phone-imgT. The optimal value is indicated in red, and the suboptimal value is indicated in blue. From the results in Tables 5 and 6, our method exhibits superior performance in terms of PSNR, SSIM, and NIQE metrics.

Table 5. A quantitative comparison of the LOL datasets using SSIM $\uparrow$, PSNR $\uparrow$, and NIQE $\downarrow$ indicators.

Metrics	PSNR $\uparrow$	SSIM $\uparrow$	NIQE $\downarrow$
RetinexNet	16.774	0.649	8.5240
GladNet	19.118	0.812	6.5391
KinD	17.648	0.825	4.9479
Zero-DCE	14.861	0.707	7.8211
LLFlow	20.998	0.835	5.7327
DRLIE	17.167	0.829	4.9848
URetinex	17.278	0.710	4.3274
Our	19.461	0.868	3.9639

Note: red is the optimal value and blue is the suboptimal value.

Table 6. A quantitative comparison of VV, DICM, MEF, and LLIV-Phone-imgT datasets using the NIQE↓ index.

Metrics	RetinexNet	GladNet	KinD	Zero-DCE
VV-Girl	3.9877	3.3858	3.2652	3.5838
VV-Women	3.7303	3.2472	2.8950	3.3505
DICM-Factory	3.5245	3.0554	3.2078	3.3119
MEF-House	4.8760	3.7013	3.5236	3.2645
LLIV-Xiaomi Mi 9	6.4722	5.0816	3.5295	5.9381
LLIV-Oppo R17	4.6039	3.8328	4.0457	4.3333
Metrics	LLFlow	DRLIE	URetinex	Our
VV-Girl	3.7142	4.5473	7.3672	2.8656
VV-Women	3.2432	3.3482	7.3040	2.6181
DICM-Factory	2.7622	3.9320	3.1259	2.9614
MEF-House	3.0593	2.7757	2.9554	3.1242
LLIV-Xiaomi Mi 9	5.6132	8.1151	5.9904	3.6709
LLIV-Oppo R17	4.2946	4.4758	4.0551	3.6255

Note: red is the optimal value and blue is the suboptimal value.

The VV dataset consists of 24 images with extremely uneven illumination, all with a size of 2304×1728 pixels. Table 7 lists the average processing time of different methods on the VV dataset on GPU. Through observation, it is evident that Zero-DCE, RetinexNet, GladNet, and DRLIE exhibit relatively shorter processing times, while our method and KinD require longer processing times. URetinex and LLFlow produce very clear enhancement results, but they take longer than our method. Considering both the processing results and processing times, our method is best.

Table 7. The comparison of the average processing times of different algorithms in the VV dataset (unit: second).

Metrics	RetinexNet	GladNet	KinD	Zero-DCE
Processing Time	0.7368	0.7925	1.1304	0.6195
Metrics	LLFlow	DRLIE	URetinex	Our
Processing Time	131.9530	0.7925	1.8530	1.2093

5. Summary

A novel low-illumination image enhancement method named TSRNet and based on Retinex is presented in this paper. In the decomposition network and the restoration network, incorporating a U-shaped network to execute trans-scale image perception functionalities enables a more accurate image decomposition and fine-grained denoising. During this process, using CBAM and the color saturation loss function, clear details and the natural color of the image are present. Furthermore, the illumination adjustment factor we proposed can adjust different brightness values, which enhances the brightness of dark regions, while avoiding the over-enhancement of bright regions. Experimental results show that the proposed method can efficiently enhance images captured under uneven and extremely low illumination. The brightness, contrast, color, and details of the resulting images exhibit notable improvements when compared to the results of existing methods. Nevertheless, the processing speed of TSRNet is relatively slow. Thus, further research should be carried out in the future.

Author Contributions: Writing—original draft preparation, Y.M.; writing—review and editing, Q.M. and X.W.; funding acquisition, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The National High-tech Research and Development Program (the 863 Program) 2022YFB3304401.

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request. You can also use this link to download the relevant data and code: <https://github.com/509-Lab/TSRNet>.

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

References

- Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **2016**, *26*, 982–993. [CrossRef] [PubMed]
- Zhao, X.; Sun, B. An attention mechanism and contextual information based low-light image enhancement method. *Int. J. Image Graph.* **2022**, *27*, 1565–1576.
- Ibrahim, H.; Kong, N.S.P. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consum. Electron.* **2007**, *53*, 1752–1758. [CrossRef]
- Abdullah-Al-Wadud, M.; Kabir, M.H.; Dewan, M.A.A. A dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consum. Electron.* **2007**, *53*, 593–600. [CrossRef]
- Li, M.; Liu, J.; Yang, W. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Trans. Image Process.* **2018**, *27*, 2828–2841. [CrossRef] [PubMed]
- Gu, Z.; Li, F.; Fang, F. A novel retinex-based fractional-order variational model for images with severely low light. *IEEE Trans. Image Process.* **2019**, *29*, 3239–3253. [CrossRef] [PubMed]
- Park, S.; Yu, S.; Moon, B. Low-light image enhancement using variational optimization-based retinex model. *IEEE Trans. Consum. Electron.* **2017**, *63*, 178–184. [CrossRef]
- Pisano, E.D.; Zong, S.; Hemminger, B.M. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *J. Digit. Imaging* **1998**, *11*, 193–200. [CrossRef]
- Shin, Y.; Jeong, S.; Lee, S. Content awareness-based color image enhancement. In Proceedings of the 18th IEEE International Symposium on Consumer Electronics, Jeju, Republic of Korea, 22–25 June 2014; pp. 1–2.
- Yelmanov, S.; Hranovska, O.; Romanyshyn, Y. A new approach to the implementation of histogram equalization in image processing. In Proceedings of the 2019 3rd International Conference on Advanced Information and Communications Technologies, Lviv, Ukraine, 2–6 July 2019; pp. 288–293.
- Jobson, D.; Rahman, Z. 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]
- Wei, C.; Wang, W.; Yang, W. Deep retinex decomposition for lowlight enhancement. In Proceedings of the British Machine Vision Conference, Newcastle, UK, 3–6 September 2018; pp. 2, 5–8.
- Wang, W.; Wei, C.; Yang, W. Gladnet: Low-light enhancement network with global awareness. In Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition, Xi'an, China, 15–19 May 2018; pp. 751–755.
- Land, E.H. The retinex theory of color vision. *Sci. Am.* **1997**, *237*, 108–129. [CrossRef]
- Li, C.; Guo, C.; Han, L. Low-light image and video enhancement using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 9396–9416. [CrossRef] [PubMed]
- 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]
- Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising with blockmatching and 3D filtering. In Proceedings of the Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning, San Jose, CA, USA, 16–18 January 2006; pp. 1–2.
- 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.
- Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18. Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
- Li, X.; Sun, Y.; Zhang, J. Low-light image enhancement via deep Retinex decomposition and bilateral learning. *Signal Process. Image Commun.* **2021**, *99*, 116466. [CrossRef]
- Zhao, Z.; Xiong, B.; Wang, L. RetinexDIP: A unified deep framework for low-light image enhancement. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 1076–1088. [CrossRef]
- He, K.; Zhang, X.; Ren, S. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- Woo, S.; Park, J.; Lee, J.Y. Cbam: Convolutional block attention module. In Proceedings of the ECCV, Munich, Germany, 8–14 September 2018; pp. 3–19.
- Zhou, Z.; Siddiquee, M.; Tajbakhsh, N. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Cham, Switzerland, 2018; pp. 3–11.
- Szegedy, C.; Vanhoucke, V.; Ioffe, S. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.

26. Ignatov, A.; Kobyshev, N.; Timofte, R. Dslr-quality photos on mobile devices with deep convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3277–3285.
27. Mittal, A.; Soundararajan, R.; Bovik, A. Making a Completely Blind Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. [[CrossRef](#)]
28. Vonikakis, V.; Kouskouridas, R.; Gasteratos, A. On the evaluation of illumination compensation algorithms. *Multimed. Tools Appl.* **2018**, *77*, 9211–9231. [[CrossRef](#)]
29. Fu, X.; Zeng, D.; Yue, H. A fusion-based enhancing method for weakly illuminated images. *Signal Process.* **2016**, *129*, 82–96. [[CrossRef](#)]
30. Lee, C.; Kim, C. Contrast enhancement based on layered difference representation. In Proceedings of the 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; pp. 965–968.
31. Boer, J.; Cense, B.; Park, B. Improved signal-to-noise ratio in spectraldomain compared with time-domain optical coherence tomography. *Opt. Lett.* **2003**, *28*, 2067–2069. [[CrossRef](#)]
32. Wang, Z.; Bovik, A.C.; Sheikh, H.R. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2014**, *13*, 600–612. [[CrossRef](#)]
33. Guo, C.; Li, C.; Guo, J. Zero-reference deep curve estimation for lowlight 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.
34. Wang, Y.; Wan, R.; Yang, W. Low-light image enhancement with normalizing flow. *AAAI Conf. Artif. Intell.* **2022**, *36*, 2604–2612. [[CrossRef](#)]
35. Tang, L.; Ma, J.; Zhang, H. DRLIE: Flexible Low-Light Image Enhancement via Disentangled Representations. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *35*, 2694–2707. [[CrossRef](#)]
36. 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.

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.