

Article

Low-Light Image Enhancement Using Photometric Alignment with Hierarchy Pyramid Network

Jing Ye, Xintao Chen, Changzhen Qiu and Zhiyong Zhang *

School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen 518107, China

* Correspondence: zhangzhy99@mail.sysu.edu.cn

Abstract: Low-light image enhancement can effectively assist high-level vision tasks that often fail in poor illumination conditions. Most previous data-driven methods, however, implemented enhancement directly from severely degraded low-light images that may provide undesirable enhancement results, including blurred detail, intensive noise, and distorted color. In this paper, inspired by a coarse-to-fine strategy, we propose an end-to-end image-level alignment with pixel-wise perceptual information enhancement pipeline for low-light image enhancement. A coarse adaptive global photometric alignment sub-network is constructed to reduce style differences, which facilitates improving illumination and revealing under-exposure area information. After the learned aligned image, a hierarchy pyramid enhancement sub-network is used to optimize image quality, which helps to remove amplified noise and enhance the local detail of low-light images. We also propose a multi-residual cascade attention block (MRCAB) that involves channel split and concatenation strategy, polarized self-attention mechanism, which leads to high-resolution reconstruction images in perceptual quality. Extensive experiments have demonstrated the effectiveness of our method on various datasets and significantly outperformed other state-of-the-art methods in detail and color reproduction.

Keywords: low-light image enhancement; coarse-to-fine; photometric alignment



Citation: Ye, J.; Chen, X.; Qiu, C.; Zhang, Z. Low-Light Image Enhancement Using Photometric Alignment with Hierarchy Pyramid Network. *Sensors* **2022**, *22*, 6799. <https://doi.org/10.3390/s22186799>

Academic Editors: Yun Zhang, KWONG Tak Wu Sam, Xu Long and Tiesong Zhao

Received: 20 August 2022
Accepted: 5 September 2022
Published: 8 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

The presence of low-light images in high-level vision tasks is inevitable, and image enhancement has a significant effect on performance improvement. However, low-light images captured in dim environments and back-lit conditions often suffer from severe degradation, including poor visibility, intensive noise, and biased color. Although long exposure shooting allows the photosensitive sensor to receive more light, which can improve the illumination of the images to a certain extent, it is impractical for real-time demanding tasks such as autonomous driving and target tracking. Due to the limitations of the camera device sensor hardware, the use of algorithms to mitigate low-light image degradation has become a research hotspot. Low-light image enhancement is mainly aimed at improving the visibility of images, removing noise, and enhancing contrast to achieve pleasant human perception effects. In the past decades, a large number of low-light image enhancement algorithms have been proposed, which can be broadly classified into the following three types: global adjustment methods, Retinex-based methods, and learning-based methods.

The main global adjustment methods include histogram equalization (HE) and gamma correction (GC). Early HE methods [1,2] enhanced the contrast by stretching the dynamic range of the images according to the histogram, and some variants of this method have been developed [3–5]. Still, the strict requirement of histogram uniform distribution severely limits the enhancement performance of such methods. The GC methods [6] adjust the values of the pixel points through an exponential function, which works on individual pixel points and ignores the relationship between adjacent pixel points, leading inevitably to over-exposure and noise amplification in the enhanced image.

Based on the traditional Retinex theory [7], the initial Retinex-based methods [8–10] estimated the illuminance and reflectance maps of low-light images and enhanced these two

components separately before fusing the output results. In recent years, several improved Retinex methods [11–14] also have been proposed to better decompose the illuminance and reflectance components by imposing prior knowledge. In [15,16], an estimated noise map was integrated into robust Retinex model to remove noise and achieve low-light enhancement.

Due to the powerful inference capabilities of machine learning techniques, learning-based methods started to develop rapidly in the field of low-light image enhancement. Benefiting from the availability of real-world paired low-/normal-light image datasets, massive methods [17–23] combine Retinex theory with deep networks, learning to estimate potential components, adjust the illumination map, and alleviate the degradations of reflectance layer for achieving natural low-light image enhancement. Yang et al. [24] proposed a band representation-based semi-supervised method to restore signal fidelity and perceptual quality. In [25–27], deep networks are constructed to generate and discriminate the high-visual-quality images; these methods release the limitations of paired datasets, effectively avoid model overfitting, and improve the generalization performance on real datasets.

Although existing learning-based methods can achieve good performance in some cases, there are still some general issues. Most of the models generate underexposed and overexposed images, lose texture and detail information during enhancement, and unpaired data training often fails to cope with distorted color and amplified noise. Simultaneously improving illumination, denoising, and restoring natural color is a non-trivial problem [28]. To address these challenges, this paper proposes an end-to-end image-level alignment with pixel-wise perceptual information enhancement pipeline for low-light image enhancement. The key insight is to minimize the style differences [29] between input low-light images and target images using an image-level alignment strategy in the coarse stage, to recover visually pleasing results at the refinement stage. Specifically, different from existing global photometric alignment methods [29] that require complicated histogram matching and gamma correction of the source domain image set, we elaborately devise a style-consistency loss to facilitate supervised learning of a global photometric alignment sub-network, which is beneficial for the adaptive style transfer of low-light images. As shown in Figure 1b, we minimize the style differences [29] (e.g., exposure, contrast, lighting, object shape, and surface textures). In the refinement stage, we develop a hierarchical pyramid enhancement sub-network to remove the amplified noise, optimize the local detail, and restore the vivid color of images; an example is given in Figure 1c. Additionally, to avoid generating artifacts and other degradations, we also design a multi-residual cascaded attention block (MRCAB), which facilitates multi-scale feature extraction and high-resolution reconstruction. The main contributions are summarized as follows:

- We propose a novel coarse-to-fine adaptive low-light image enhancement network (CFANet) that seamlessly combines coarse global photometric alignment with finer perceptual information promotion. The coarse-to-fine pipeline is trained in a data-driven manner within a unified framework to avoid error accumulation.
- The built MRCAB is embedded into a hierarchy pyramid network, which can change the perceptual fields and highlight notable features for each network layer. Furthermore, the polarized self-attention mechanism of the block can preserve high-resolution information to achieve better enhancement performance.
- Experiments show that our method can generalize well across different real low-light datasets. Specially, we restore less noise normal-light images with rich detail and vivid colors compared to other low-light enhancement methods.



Figure 1. Examples of low-light image enhancement using our method. (a) Input low-light images. (b) The aligned images resulting from adaptive global photometric alignment improve the contrast and reveal underexposed regions. (c) The final results display rich detail and vibrant color.

2. Related Work

2.1. Traditional Methods

Traditional methods mainly review global adjustment methods and Retinex-based methods. The classical global HE method [1,2] implements nonlinear stretching to enhance image contrast and reveal the content of underexposed areas, but it may cause overexposure and drowning of detail by over-transforming the saturated regions. To cope with this problem, in the local HE method [3], the global histogram is sliced into multiple sub-histograms and the enhancement operations are performed separately in different regions, which helps to improve the performance of low-light image enhancement flexibly. However, these methods increase the computational complexity to some extent. Therefore, the parametric-oriented HE methods [4,5] attempt to reduce the enhancement complexity by optimizing the transformation process to a uniform function that maps the low-light images to the output results. Huang et al. [6] improved the contrast of images by gamma correction of luminance pixels. However, the above methods are not designed for low-light image enhancement especially, and the enhancement results often show hard noise and unnatural results.

Single-scale Retinex [8] is the first practical application of Retinex theory to image processing, and it is found that the surround formation produces the best enhancement results. Single-scale Retinex was extended to Multi-scale Retinex [9] to achieve both color and luminance recovery. Some studies [11–13] decomposed the illumination and reflectance components by using artificially designed priors, ignoring the degradations of the reflectance layer, which may lead to strong noise in the output. On the other hand, a noise prior was added to constructing a robust Retinex model for enhancing low-illumination images in [15,16]. The prior design of the above methods is too complex and cannot satisfy the adaptive enhancement of the real low-light images.

2.2. Learning-Based Methods

Learning-based methods have achieved extraordinary results in several vision domains. Loree et al. [30] was the first to explore the application of deep learning in the field of low-light image enhancement and proposed the deep autoencoder-based method (LLNet), which obtained impressive enhancement results. In [17,31], LOL and SID real low-/normal-light pairs were proposed to accelerate the development of learning-based low-light enhancement methods. MBLEN [32] uses multi-branch sub-networks to enhance the different layer inputs separately, then outputs the fusion results. Zhang et al. [18] constructed three sub-networks for decomposition, adjustment of illumination, and recovery of reflectance components, respectively. In [22], KinD++ was proposed to mitigate visual defects (non-uniform spots and over-smoothing) left in KinD [18]. Lu et al. [33] proposed slight and heavy adaptive attention mechanisms for low-light image enhancement with different degrees of degradations. Li et al. [34] used a luminance-aware pyramidal structure to enhance the local and global features of the low-light images. These works focus on enhancing severely degraded low-light images directly via improving the network structure. However, simultaneously boosting illumination, removing noise, and restoring detail can lead to undesirable results. Some later methods adopted more special deep networks than the previous works. Li et al. [35,36] used the deep curve estimation to achieve impressive results. Jiang et al. [25] first introduced an unpaired learning strategy to build a new pipeline EnlightenGAN, which greatly improved the generalization performance of the model. Pan et al. [26] proposed a multi-module cascade generative network and adaptive multi-scale discriminative network. However, unsupervised methods lack the guidance of paired data and need to be further improved in terms of image fidelity and color recovery ability.

Comparatively, this paper enhances low-light images in a coarse-to-fine manner. Inspired by the image-level domain shift strategy [29], an adaptive global photometric alignment sub-network is used to shift the style of severe degraded low-light images, including exposure, contrast, and texture, with the ability to explore the content in underexposed regions. In the optimization stage, local detail, and color of the aligned image are further enhanced to remove the amplified noise and generate visually pleasurable images.

3. Coarse-to-Fine Enhancement Pipeline

3.1. Motivation

Deep methods can effectively enhance the quality of low-light images. In general, performing enhancement tasks directly from severely degraded low-light images usually yield undesirable results. In other words, simultaneously enhancing illumination, removing noise terms, and restoring vivid color would also be a very difficult task.

Is it possible to perform low-light image enhancement progressively? The methods in [33,37] support this view. However, the former requires stepwise training of two independent networks, which is prone to the accumulation of model errors. Additionally, the latter lacks the guidance of paired data, and there are amplified noise and distorted colors in the recovered results. Based on the above observations, it is feasible to effectively enhance low-light images in a coarse-to-fine manner, and the enhancement work is facilitated by obtaining an intermediate image from the input image that is close to the target image in terms of brightness, contrast, and content. This is essentially different from the coarse-to-fine network framework used in [34,38] for extracting multi-scale features. Furthermore, a recently proposed domain shifts problem [29] inspired us. In their work, the image-level alignment is used to decrease the domain shifts. Based on the success of the semantic segmentation task, the idea of estimating photometric aligned images motivated us to extend it to style transformation in low-light image enhancement. However, performing classic histogram matching on color channels and lightness gamma correction on the light channel ignores the association between the channels and cannot accommodate diverse low-light image enhancement.

Based on the above insights, our CFANet attempts to enhance low-light images at the image-level and pixel-wise in a coarse-to-fine fashion, adaptive style transfer of low-light

images is implemented using a deep model, which better simulates the enhancement of low-light images and works effectively within a unified network framework. Particularly, the intermediate style consistency loss can better boost brightness and explore content in underexposed areas. This unique design allows CFANet to overcome the problem of independent channel adjustment. Hence, it trades off well between style difference and content preservation.

3.2. CFANet

Figure 2 shows the overall architecture of CFANet, which can be divided into two sub-networks. The coarse adaptive global photometric alignment sub-network learns the style transformation of low-light images, and the finer hierarchy pyramid enhancement sub-network uses multi-residual cascade attention blocks (MRCABs) to further optimize the aligned images. We described the two sub-networks and MRCAB in detail below.

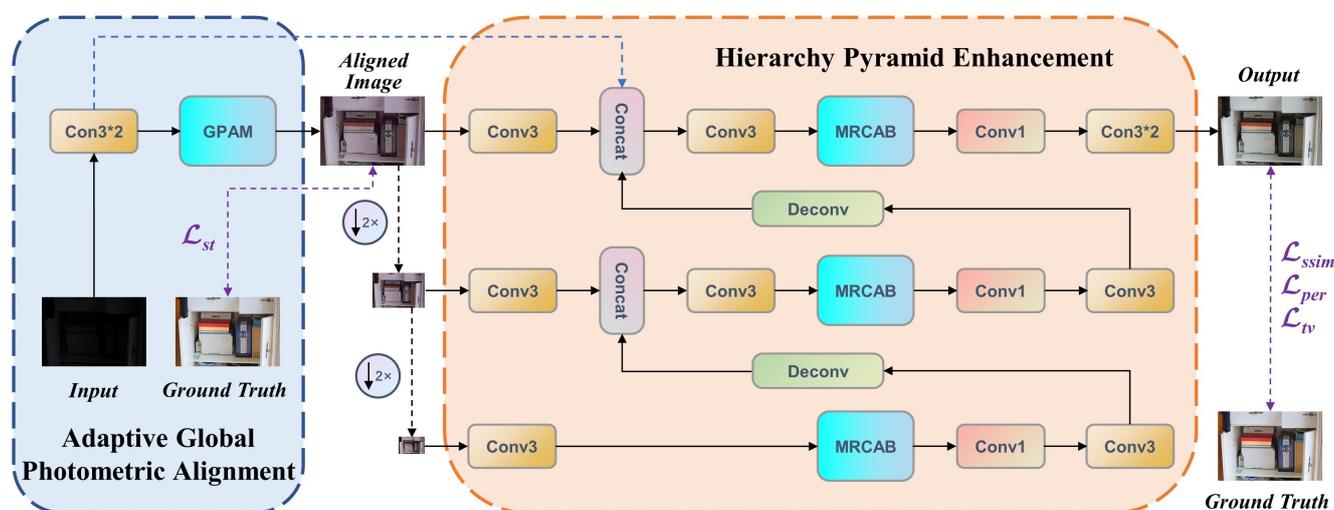


Figure 2. An illustration of the proposed CFANet. It contains an adaptive global photometric alignment sub-network for style transformation, and a hierarchy pyramid enhancement sub-network for optimization of the image quality. We build the GPAM with the basic U-net structure. The proposed MRCABs are inserted in hierarchy pyramid architecture to extract multi-scale features in a wider range, suppressing artifacts and color distortion more efficiently. The low-light images are mapped to the output in a coarse-to-fine manner.

3.2.1. Network Architecture

In the coarse stage, the input low-light images are fed into an adaptive global photometric alignment sub-network that is designed to decrease style differences with the supervision of style consistency loss \mathcal{L}_{st} (see Section 3.3). Therefore, given a collection of M image pairs $\{I_{in}^m, I_{gt}^m\}_{m=1}^M$, we aim to solve the following problem:

$$\gamma^* = \arg \min_{\gamma} \frac{1}{M} \sum_{m=1}^M \mathcal{L}_{st} \left(E_{AGPA_{\gamma}}(I_{in}^m), I_{gt}^m \right), \quad (1)$$

where I_{in}^m and I_{gt}^m denote the input image and ground truth, respectively. γ is the parameter set and $E_{AGPA}(\cdot)$ represents the adaptive global photometric alignment sub-network. Here, $\mathcal{L}_{st}(\cdot)$ is adopted to minimize the style difference between the aligned image and ground truth. The coarse network consists of two 3×3 convolutional layers and a global photometric alignment module (GPAM), as shown in Figure 2. Two convolutional layers in the front-end of the network are first used to explore shallow features of low-light images. After that, we build the GPAM with the basic U-net [39] structure, [25,36] also demonstrated the effectiveness of the U-net in low-light image enhancement. Thanks to the

skip connections between the downsampling and upsampling layers of the GPAM, the sub-network can preserve the structure information of the original images while enhancing the brightness, contrast, and exploring the content of underexposed regions during the style transformation. To facilitate intermediate supervision, we output an aligned image by the last convolution layer. Our design gears adaptive global photometric alignment sub-network to embed the input low-light images into the feature space of aligned images, allowing the subsequent hierarchy pyramid enhancement sub-network to pay more attention to optimization tasks.

Although the aligned images from the adaptive global photometric alignment sub-network are close to the target images in terms of luminance, contrast, and surface texture. However, as can be observed in Figure 1b, there are still color distortions, artifacts, and amplified noise. In the refinement stage, the hierarchy pyramid enhancement sub-network focuses on optimizing the above problems. Essentially, this sub-network also enhances the features in a coarse-to-fine strategy. The aligned images are used as input images of different resolutions after downsampling. Although different branches consist of MRCAB with the same structure, the multi-scale network can enhance global and local features from the bottom up, respectively. Furthermore, to avoid the effect of detail information loss caused by over-convolution, after all global features are pooled to the top branch by the deconvolution operation, a skip connection is established to share shallow features to improve the refinement features and generate the final results.

In particular, it is important to note that our CFANet is implemented in a data-driven manner within a unified framework, which is beneficial for decreasing error accumulation and restoring desirable normal-light images.

3.2.2. Multi-Residual Cascade Attention Block (MRCAB)

Both photometric alignment and perceptual quality improvement in our task are spatially varying problems. Though the hierarchy pyramid architecture can explore features at different scales, it is not enough for image quality optimization tasks. Typical low-light image enhancement networks are prone to artifacts and unnatural color, and we find that these problems can be significantly remedied by changing the perceptual field of the network, highlighting and suppressing features. To achieve this goal, we elaborately devise the MRCAB, which consists of four cascaded Res2Net [40] and a polarized self-attention (PSA) block [41] (see Section 5 for a detailed description of the cascade number settings of Res2Net), as shown in Figure 3. The Res2Net adopts channel split and concatenation strategy to form different receptive fields to effectively extract multi-scale features. Rather than using the SE block [42], we choose a PSA block that is more adapted to pixel-wise regression, the attention mechanism simultaneously maintains high resolution in both channel and space, and achieves nonlinear enhancement of high-resolution information. Specifically, skip connections within MRCAB allow for more efficient utilization and propagation of hierarchical feature information. We would like to highlight that MRCAB is the essential component in our network, it suppresses artifacts and color distortion that may be caused by the co-existence of other degradations of oversaturation and noise.

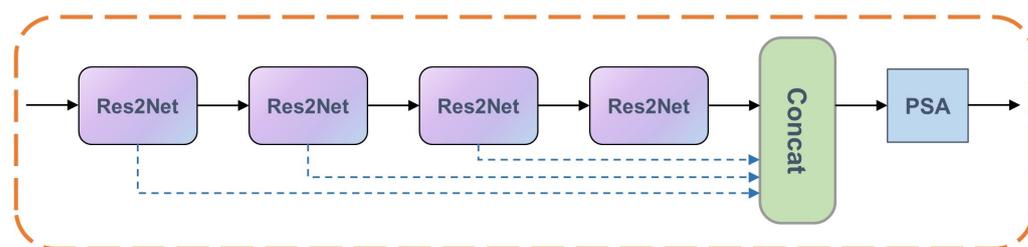


Figure 3. An illustration of our proposed multi-residual cascade attention block(MRCAB), which is the crucial component of the CFANet. It consists of four cascaded Res2Net and a polarized self-attention (PSA) block.

3.3. Loss Function

To enable supervised learning in the coarse-to-fine pipeline, our proposed loss function consists of the following four parts.

Style consistency loss. To reduce the style difference between the aligned image and the ground truth, boost the brightness and preserve the detail of the original image. Specifically, we provide intermediate supervision at the end of the adaptive global photometric alignment sub-network. The style consistency loss \mathcal{L}_{st} can be expressed as:

$$\mathcal{L}_{st} = \sum_{c \in \xi} |J_a^c - J_g^c|, \xi = \{R, G, B\}, \quad (2)$$

where J_a^c, J_g^c denote the intensity value of the aligned image and the ground truth in channel c , respectively.

Structure similarity loss. Since MAE and MSE losses ignore the correlation between the long-distance of pixel points, this makes it difficult to overcome structural distortions such as artifacts and blurring. Therefore, we introduce the structure similarity loss to enhance the recovery quality of low-light images. The structure similarity loss \mathcal{L}_{ssim} is defined as:

$$\mathcal{L}_{ssim} = 1 - \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \cdot \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}, \quad (3)$$

where μ_x, μ_y represent the pixel average value of x, y images, respectively. σ_x^2 and σ_y^2 are variances, the covariance is represented as σ_{xy} . To avoid the denominator being zero, c_1, c_2 are set to 0.0001 and 0.0009 in our work, following the same setting in Wang et al. [43].

Perceptual loss. To facilitate the enhancement of image perceptual information, we fed the enhanced image and the ground truth into the pre-trained VGG-19 network to measure the difference between the corresponding feature maps. The perceptual loss \mathcal{L}_{per} can be expressed as:

$$\mathcal{L}_{per} = \frac{1}{C_n H_n W_n} \|\phi_n(E(I_{in})) - \phi_n(I_{gt})\|_2^2, \quad (4)$$

where $E(\cdot)$ represents our CFANet. $\phi_n(\cdot)$ is the feature map of the n -th convolutional layer in VGG-19 model. C_n, H_n, W_n denote the dimensions of the corresponding feature maps, respectively.

Total variation loss. To remove noise and improve the visual effect of the images, total variation loss is introduced to limit the gradient of the images. The total variation loss \mathcal{L}_{tv} is written as:

$$\mathcal{L}_{tv} = \sum_{i=1}^H \sum_{j=1}^W \sqrt{(p_{ij} - p_{i+1,j})(p_{ij} - p_{i,j+1})}, \quad (5)$$

where p represents the intensity value at pixel point index (i, j) .

Total loss. The total loss function is:

$$\mathcal{L}_{total} = \lambda_{st}\mathcal{L}_{st} + \mathcal{L}_{ssim} + \lambda_{per}\mathcal{L}_{per} + \lambda_{tv}\mathcal{L}_{tv}. \quad (6)$$

We set the loss weights of λ_{st} , λ_{per} , and λ_{tv} to 0.1, 0.2, and 0.01, respectively, in our experiments.

4. Experiments

4.1. Datasets and Evaluate Metrics

We train our CFANet and other state-of-the-art methods using LOL [17] and SID [31] datasets. The LOL dataset consists of 500 real scenes image pairs and 1000 synthetic image pairs, the SID dataset in RAW format is converted to sRGB format for training. Additionally, we also evaluated on LIME [12], MEF [44], NPE [10], DICM [5], VV [45]

datasets to demonstrate the effectiveness and generality of our approach. We adopt the commonly used PSNR, SSIM [46], and NIQE [47] metrics for evaluation.

4.2. Experimental Settings

The proposed CFANet is designed based on the Pytorch framework. We randomly crop 256×256 patches for training on NVIDIA RTX 3090 GPU, all these patches are transformed by randomly flipping and rotations of 90° , 180° , 270° . The network is trained on the LOL and SID datasets for 400,300 epochs, respectively. The former has an initial learning rate of 10^{-4} , which is halved at 200 epochs; and the latter has an initial learning rate of 10^{-3} , which is halved at 150 epochs. The mini-batch is set to 8. We train our network using Adam optimizer with $\beta_1 = 0.9$; $\beta_2 = 0.99$.

4.3. Enhancement Results

To comprehensively evaluate the low-light image enhancement performance of CFANet, we performed quantitative evaluations on LOL, SID, LIME, MEF, NPE, DICM, VV datasets, and qualitative comparisons on datasets besides SID.

4.3.1. Quantitative Evaluation

We choose recent light enhancement networks to evaluate the performance of LOL synthetic and real datasets, which is consistent with the evaluation approach in [21,24], including BIMEF [48], CRM [49], DHECE [50], Dong [51], EFF [52], LIME [12], MF [11], MBLLN [32], JED [16], SRIE [13], RRM [15], DRD [17], DeepUPE [53], SCIE [54], KinD [18], EnlightenGAN [25], RetinexNet [21], KinD++ [22], and DRBN [24]. As shown in Table 1, we found both on synthetic and real LOL datasets that our method achieves the best results in both PSNR and SSIM metrics compared to the state-of-the-art methods. The results suggest that CFANet is effective and particularly well-suited for low-light image enhancement tasks.

Table 1. Quantitative evaluation on LOL synthetic and real dataset, in terms of PSNR and SSIM. The best results are in bold.

Method	LOL-Syn		LOL-Real	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
BIMEF	17.20	0.7172	17.85	0.6526
CRM	18.91	0.7864	19.65	0.6623
DHECE	17.75	0.7800	14.64	0.4450
Dong	16.90	0.7487	17.26	0.5270
EFF	17.20	0.7127	17.85	0.6526
LIME	16.88	0.7762	15.24	0.4702
MF	17.50	0.7514	18.73	0.5590
MBLLN	17.07	0.7301	17.86	0.7247
JED	17.48	0.7444	17.33	0.6654
SRIE	14.50	0.6163	17.34	0.6859
RRM	17.15	0.7277	17.33	0.5144
DRD	17.13	0.7978	15.47	0.5672
DeepUPE	15.08	0.6225	13.27	0.4521
SCIE	18.50	0.7631	19.40	0.6906
KinD	17.84	0.7971	20.73	0.8103
EnlightenGAN	16.57	0.7338	18.23	0.6165
RetinexNet	22.05	0.9054	20.06	0.8158
KinD++	17.69	0.8334	21.30	0.8226
DRBN	23.22	0.9275	20.29	0.8310
Our	24.62	0.9314	21.64	0.8481

Since linear RAW data is significantly different from nonlinear sRGB data, the model trained in RAW format cannot be adapted to enhance sRGB images, and the image format acquired by photographic devices is usually sRGB [55]. Therefore, this paper only compares with networks trained on the SID dataset in sRGB format, including DSLR [56], LIME [12], SCIE [54], DeepUPE [53], and LRD [55]. For the test results of the SID dataset in Table 2, we found that our network achieved the best results in the PSNR metric and comparable results in the SSIM metric, which shows the superiority of our coarse-to-fine strategy and losses.

Table 2. Quantitative evaluation on SID dataset, in terms of PSNR and SSIM. The best results are in bold.

Method	PSNR \uparrow	SSIM \uparrow
DSLR	17.25	0.4229
LIME	17.76	0.3506
SCIE	21.16	0.6398
DeepUPE	21.55	0.6531
LRD	22.13	0.7172
Our	22.60	0.6728

We evaluated the proposed CFANet and nine representative methods on several real datasets LIME, MEF, NPE, DICM, and VV. Table 3 shows the NIQE metric test results, no single method can achieve the best score on all datasets, but our method performs the best on NPE and DICM datasets, and otherwise still maintains a good score on other datasets. The comparisons in real datasets strongly suggest the effectiveness and generality of our proposed network.

Table 3. Quantitative evaluation with NIQE metric on LIME, MEF, NPE, DICM, and VV datasets. The best results are in bold.

Method	NIQE \downarrow				
	LIME	MEF	NPE	DICM	VV
SRIE	3.3481	3.1601	3.6930	3.4161	3.0015
LIME	3.3176	2.9363	3.9679	3.5289	2.4221
RRM	4.1056	4.3742	4.3785	3.9799	4.3785
MBLLEN	4.1445	4.1969	4.2200	4.0426	4.0631
KinD	4.5086	3.3126	3.7476	3.8037	2.9148
DeepUPE	3.6233	3.4051	4.0390	3.9296	3.1807
RetinexNet	4.0272	3.9265	4.1013	4.1775	2.5792
EnlightenGAN	3.1880	2.9440	3.6775	3.3632	2.5875
KinD++	4.3394	3.3082	3.8462	3.5727	2.5974
Our	3.6706	2.9956	3.6124	3.3186	3.1158

4.3.2. Qualitative Evaluation

In this part, the results of three traditional methods and five deep learning methods in Figures 4–9 are described in detail in comparison with our network in terms of visual effects. We found that SRIE produced underexposed enhancement results in most cases (e.g., Figures 4, 6, 8 and 9) and less improvement for image contrast. LIME generated several overexposed regions in Figures 6–8, and the work adopted the denoising mechanism as post-processing still caused strong noise and artifacts. To effectively reduce the effect of noise, RRM improves the robustness by estimating the noise map in the model, but over smoothing the image causes blurring of the main structures and loss of image detail information in Figures 4, 5, 7 and 9. Figures 4, 5, 6 and 9 show that the low-light enhancement performance of DeepUPE is weak, producing a large number of unexposed areas. Early RetinexNet performed poorly in terms of enhancement performance, with sig-

nificant noise and artifacts in all enhanced images. The unsupervised methods Zero-DCE and EnlightenGAN are trained on unpaired data, they restore relatively impressive results on different datasets, but also suffer from color distortion (e.g., Figures 5 and 7) and fail to cope with extremely dark regions (e.g., Figure 4). KinD++ overcame the visual defects of excessive smoothing and uneven brightness to a certain extent by improving the KinD method; however, we found that there are still problems of unclear image detail and low contrast in Figures 4 and 7.

In comparison with the above results, we restored normal-light images of good visual quality in all enhancement experiments. Thanks to the particular frame design of CFANet, our method can explore the content of underexposed regions using the adaptive global light alignment module while maintaining high resolution. In particular, as shown in Figures 4, 5 and 7, beneficial from the coarse-to-fine strategy, the images processed through our network exhibit stunning colors and excellent contrast, with clear detail and good illumination for a pleasant visual effect. The visual comparison in various cases indicates the superiority and generalization of our approach.

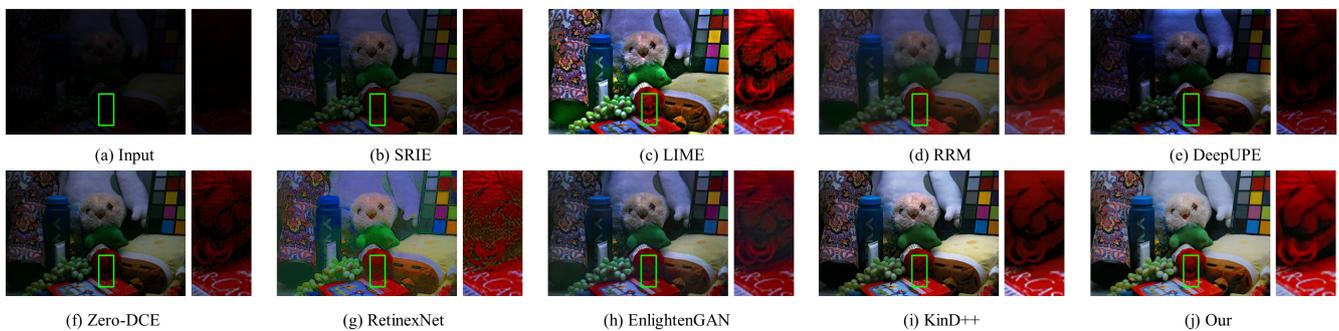


Figure 4. Visual comparison of low-light image enhancement on an image from the LOL dataset.

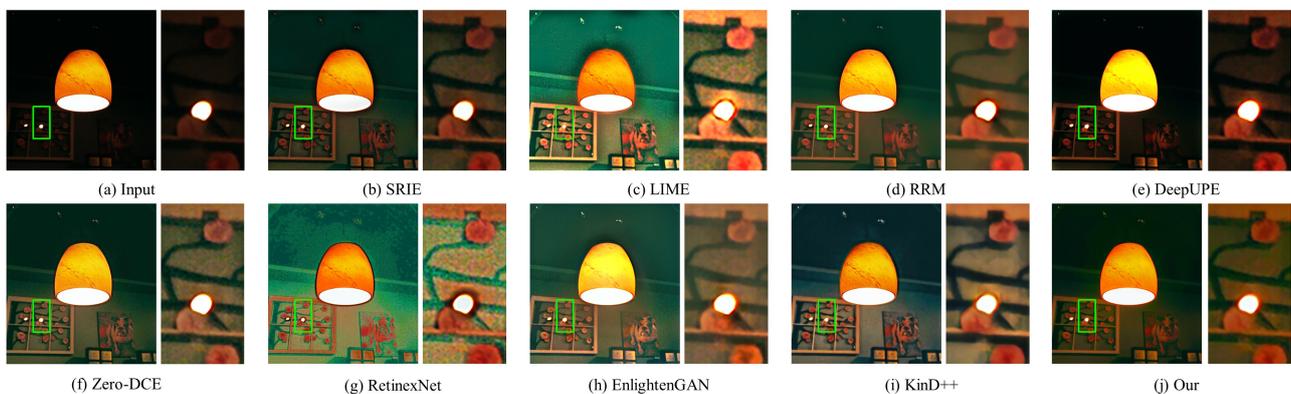


Figure 5. Visual comparison of low-light image enhancement on an image from the LIME dataset.

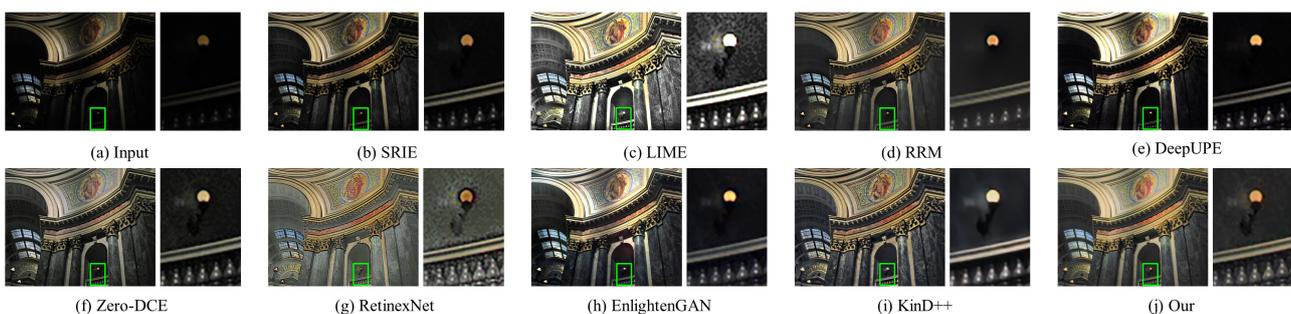


Figure 6. Visual comparison of low-light image enhancement on an image from the MEF dataset.

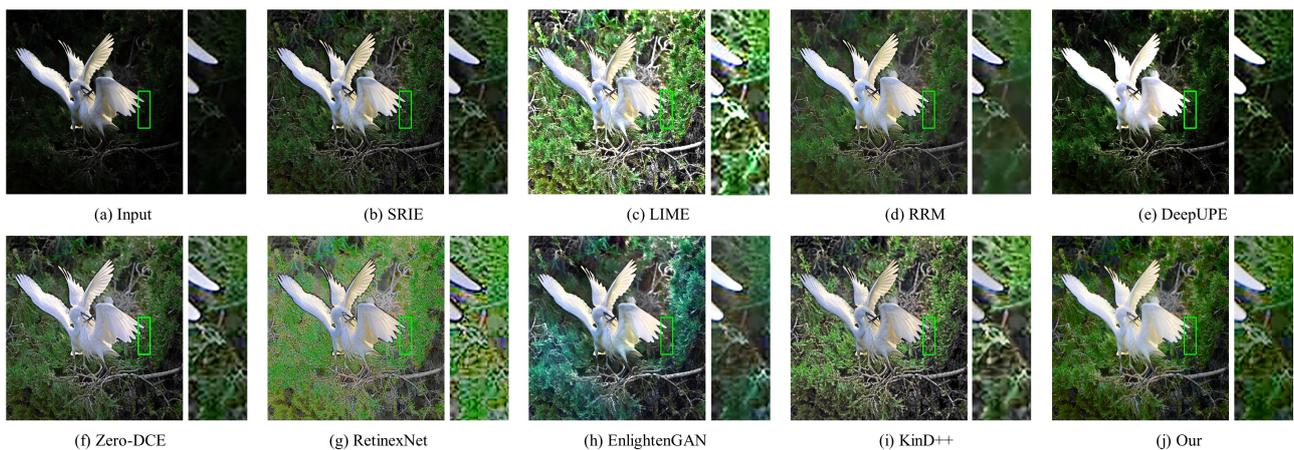


Figure 7. Visual comparison of low-light image enhancement on an image from the NPE dataset.

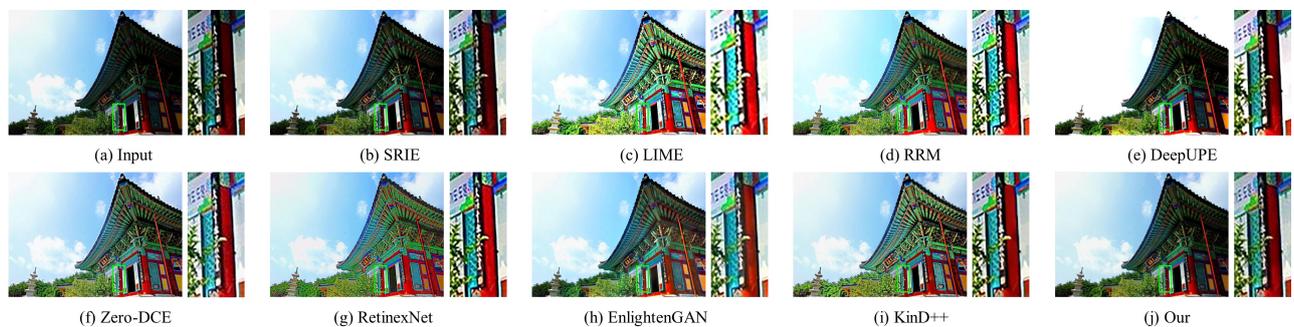


Figure 8. Visual comparison of low-light image enhancement on an image from the DICM dataset.



Figure 9. Visual comparison of low-light image enhancement on an image from the VV dataset.

Though our method achieves promising results in most cases, but we also found that the model may show fragile performance in extreme darkness, such as the artifacts in the face region of Figure 9, which is degradation caused by over-smoothing to avoid noise. To cope with this limitation, we plan to implicitly incorporate the denoising process into our model to mitigate this problem in the future.

5. Ablation Study

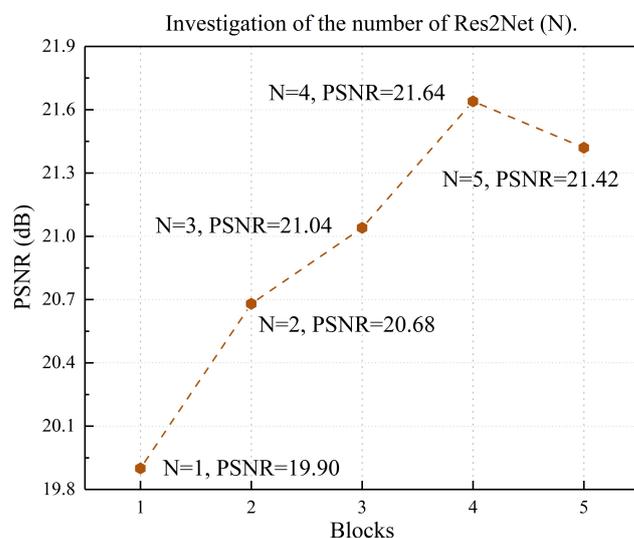
In this section, we present an ablation study to demonstrate the effectiveness of the main components in CFANet and losses, which was performed on the LOL dataset.

Effectiveness of network architecture. As shown in Table 4, in the absence of a global photometric alignment module, the Res2Net module performs slightly better than the ResNet module, and the performance is further improved by adding the PSA mechanism, while our proposed CFANet, which includes the global light alignment module, achieves the best scores on both PSNR and SSIM metrics. The above quantitative results demonstrated the effectiveness of our network components.

Table 4. Ablation results of network structure on LOL dataset. The best results are in bold.

Metric	Module			
	ResNet	Res2Net	Res2Net + PSA	CFANet
PSNR	19.72	19.94	20.83	21.64
SSIM	0.8037	0.8078	0.8143	0.8481

To investigate the effect of the number of Res2Net blocks in MRCAB on low-light image enhancement, we set different numbers of blocks to train the model. As we can see in Figure 10, with the increase of Res2Net blocks, the network gradually improves the PSNR. When the number of blocks exceeds four, the benefit of improving PSNR disappears, and the network is prone to overfitting results. We found that the optimal setting for the number of blocks is $N = 4$.

**Figure 10.** Investigate the effect of Res2Net block number for enhancement performance on the LOL dataset.

Effectiveness of losses. We verify the validity of each loss function by adding them step by step. In Table 5, removing arbitrary losses degrades the network performance. The combination of style consistency loss, structural similarity loss, perceptual loss, and total variance loss achieves the best performance, which also indicates that the intermediate style consistency loss is effective for our network.

Table 5. Ablation results of loss function on LOL dataset. The best results are in bold.

Loss Configuration	PSNR	SSIM
1. with \mathcal{L}_{ssim} , w/o \mathcal{L}_{tv} , w/o \mathcal{L}_{per} , w/o \mathcal{L}_{st}	20.19	0.8225
2. with \mathcal{L}_{ssim} , with \mathcal{L}_{tv} , w/o \mathcal{L}_{per} , w/o \mathcal{L}_{st}	21.43	0.8341
3. with \mathcal{L}_{ssim} , with \mathcal{L}_{tv} , with \mathcal{L}_{per} , w/o \mathcal{L}_{st}	21.40	0.8367
4. default configuration	21.64	0.8481

6. Conclusions

In this paper, we have presented a novel coarse-to-fine adaptive low-light image enhancement pipeline that seamlessly combined coarse global photometric alignment with finer perceptual information promotion. With the coarse adaptive global photometric alignment subnet, the difference in style between low-light and normal-light images is effectively reduced, facilitating improved illumination and revealing information in underexposed

areas. Moreover, the proposed multi-residual cascade attention block (MRCAB) is designed to be embedded in the backbone network, which allows CFANet to avoid degradations and maintain high resolution. Compared to other low-light image enhancement algorithms, our proposed CFANet achieves significant improvements in PSNR and SSIM, and restores suitable illumination, rich detail information, and vivid colors. Extensive experiments on widely used low-light image datasets have demonstrated the effectiveness and generality of our method.

Our method can effectively mitigate the detail blur of static images. In general, real-world low-light images usually have the problem of image blur caused by fast target movement and camera shake [57], we will explore solutions for the joint task of low-light image enhancement and deblurring in future work.

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

Funding: This research received no external funding.

Data Availability Statement: The dataset and the code of the comparison method used in this paper are publicly available on github.

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

References

1. Chen, S.D.; Ramli, A.R. Contrast enhancement using recursive mean-separate histogram equalization for scalable brightness preservation. *IEEE Trans. Consum. Electron.* **2003**, *49*, 1301–1309. [[CrossRef](#)]
2. Celik, T.; Tjahjadi, T. Contextual and variational contrast enhancement. *IEEE Trans. Image Process.* **2011**, *20*, 3431–3441. [[CrossRef](#)] [[PubMed](#)]
3. Celik, T.; Tjahjadi, T. Automatic image equalization and contrast enhancement using Gaussian mixture modeling. *IEEE Trans. Image Process.* **2011**, *21*, 145–156. [[CrossRef](#)] [[PubMed](#)]
4. Liu, Y.F.; Guo, J.M.; Lai, B.S.; Lee, J.D. High efficient contrast enhancement using parametric approximation. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 2444–2448.
5. 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](#)]
6. Huang, S.C.; Cheng, F.C.; Chiu, Y.S. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Trans. Image Process.* **2012**, *22*, 1032–1041. [[CrossRef](#)] [[PubMed](#)]
7. Land, E.H. The retinex theory of color vision. *Sci. Am.* **1977**, *237*, 108–129. [[CrossRef](#)]
8. Jobson, D.J.; Rahman, Z.U.; Woodell, G.A. Properties and performance of a center/surround retinex. *IEEE Trans. Image Process.* **1997**, *6*, 451–462. [[CrossRef](#)]
9. Jobson, D.J.; Rahman, Z.U.; Woodell, G.A. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.* **1997**, *6*, 965–976. [[CrossRef](#)]
10. 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](#)]
11. Fu, X.; Zeng, D.; Huang, Y.; Liao, Y.; Ding, X.; Paisley, J. A fusion-based enhancing method for weakly illuminated images. *Signal Process.* **2016**, *129*, 82–96. [[CrossRef](#)]
12. 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](#)]
13. Fu, X.; Zeng, D.; Huang, Y.; Zhang, X.P.; Ding, X. A weighted variational model for simultaneous reflectance and illumination estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2782–2790.
14. Fu, X.; Zeng, D.; Huang, Y.; Ding, X.; Zhang, X.P. A variational framework for single low light image enhancement using bright channel prior. In Proceedings of the 2013 IEEE Global Conference on Signal and Information Processing, Austin, TX, USA, 3–5 December 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 1085–1088.
15. 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](#)] [[PubMed](#)]

16. Ren, X.; Li, M.; Cheng, W.H.; Liu, J. Joint enhancement and denoising method via sequential decomposition. In Proceedings of the 2018 IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 27–30 May 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–5.
17. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep retinex decomposition for low-light enhancement. *arXiv* **2018**, arXiv:1808.04560.
18. 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.
19. Wang, Y.; Cao, Y.; Zha, Z.J.; Zhang, J.; Xiong, Z.; Zhang, W.; Wu, F. Progressive retinex: Mutually reinforced illumination-noise perception network for low-light image enhancement. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2015–2023.
20. 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, Virtual, 19–25 June 2021; pp. 10561–10570.
21. 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](#)] [[PubMed](#)]
22. Zhang, Y.; Guo, X.; Ma, J.; Liu, W.; Zhang, J. Beyond brightening low-light images. *Int. J. Comput. Vis.* **2021**, *129*, 1013–1037. [[CrossRef](#)]
23. Jiang, Z.; Li, H.; Liu, L.; Men, A.; Wang, H. A switched view of Retinex: Deep self-regularized low-light image enhancement. *Neurocomputing* **2021**, *454*, 361–372. [[CrossRef](#)]
24. Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; Liu, J. Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality. *IEEE Trans. Image Process.* **2021**, *30*, 3461–3473. [[CrossRef](#)]
25. 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](#)]
26. Pan, Z.; Yuan, F.; Lei, J.; Li, W.; Ling, N.; Kwong, S. MIEGAN: Mobile image enhancement via a multi-module cascade neural network. *IEEE Trans. Multimed.* **2021**, *24*, 519–533. [[CrossRef](#)]
27. Zhao, Z.; Xiong, B.; Wang, L.; Ou, Q.; Yu, L.; Kuang, F. RetinexDIP: A unified deep framework for low-light image enhancement. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 1076–1088. [[CrossRef](#)]
28. Hu, J.; Guo, X.; Chen, J.; Liang, G.; Deng, F.; Lam, T.L. A two-stage unsupervised approach for low light image enhancement. *IEEE Robot. Autom. Lett.* **2021**, *6*, 8363–8370. [[CrossRef](#)]
29. Ma, H.; Lin, X.; Wu, Z.; Yu, Y. Coarse-to-Fine Domain Adaptive Semantic Segmentation with Photometric Alignment and Category-Center Regularization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4051–4060.
30. Lore, K.G.; Akitayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **2017**, *61*, 650–662. [[CrossRef](#)]
31. Chen, C.; Chen, Q.; Xu, J.; Koltun, V. Learning to see in the dark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3291–3300.
32. Lv, F.; Lu, F.; Wu, J.; Lim, C. MBLEN: Low-Light Image/Video Enhancement Using CNNs. In Proceedings of the 2018 British Machine Vision Conference, Newcastle, UK, 3–6 September 2018; Volume 220, pp. 1–13.
33. Lu, K.; Zhang, L. TBEFN: A two-branch exposure-fusion network for low-light image enhancement. *IEEE Trans. Multimed.* **2020**, *23*, 4093–4105. [[CrossRef](#)]
34. Li, J.; Li, J.; Fang, F.; Li, F.; Zhang, G. Luminance-aware pyramid network for low-light image enhancement. *IEEE Trans. Multimed.* **2020**, *23*, 3153–3165. [[CrossRef](#)]
35. Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1780–1789.
36. Liu, M.; Tang, L.; Zhong, S.; Luo, H.; Peng, J. Learning noise-decoupled affine models for extreme low-light image enhancement. *Neurocomputing* **2021**, *448*, 21–29. [[CrossRef](#)]
37. Zhu, M.; Pan, P.; Chen, W.; Yang, Y. Eemefn: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In Proceedings of the AAAI Conference on Artificial Intelligence, Hilton New York Midtown, NY, USA, 7–12 February 2020; Volume 34, pp. 13106–13113.
38. Lim, S.; Kim, W. Dslr: Deep stacked laplacian restorer for low-light image enhancement. *IEEE Trans. Multimed.* **2020**, *23*, 4272–4284. [[CrossRef](#)]
39. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing And Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin, Germany, 2015; pp. 234–241.
40. Gao, S.H.; Cheng, M.M.; Zhao, K.; Zhang, X.Y.; Yang, M.H.; Torr, P. Res2net: A new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 652–662. [[CrossRef](#)]
41. Liu, H.; Liu, F.; Fan, X.; Huang, D. Polarized self-attention: Towards high-quality pixel-wise regression. *arXiv* **2021**, arXiv:2107.00782.
42. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.

43. Wang, L.W.; Liu, Z.S.; Siu, W.C.; Lun, D.P. Lightning network for low-light image enhancement. *IEEE Trans. Image Process.* **2020**, *29*, 7984–7996. [[CrossRef](#)]
44. Ma, K.; Zeng, K.; Wang, Z. Perceptual quality assessment for multi-exposure image fusion. *IEEE Trans. Image Process.* **2015**, *24*, 3345–3356. [[CrossRef](#)]
45. Vonikakis, V.; Andreadis, I.; Gasteratos, A. Fast centre-surround contrast modification. *IET Image Process.* **2008**, *2*, 19–34. [[CrossRef](#)]
46. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
47. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “completely blind” image quality analyzer. *IEEE Signal Process. Lett.* **2012**, *20*, 209–212. [[CrossRef](#)]
48. Ying, Z.; Li, G.; Gao, W. A bio-inspired multi-exposure fusion framework for low-light image enhancement. *arXiv* **2017**, arXiv:1711.00591.
49. Ying, Z.; Li, G.; Ren, Y.; Wang, R.; Wang, W. A new low-light image enhancement algorithm using camera response model. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 3015–3022.
50. Nakai, K.; Hoshi, Y.; Taguchi, A. Color image contrast enhancement method based on differential intensity/saturation gray-levels histograms. In Proceedings of the 2013 International Symposium on Intelligent Signal Processing and Communication Systems, Okinawa, Japan, 12–15 November 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 445–449.
51. Dong, X.; Wang, G.; Pang, Y.; Li, W.; Wen, J.; Meng, W.; Lu, Y. Fast efficient algorithm for enhancement of low lighting video. In Proceedings of the 2011 IEEE International Conference on Multimedia and Expo, Barcelona, Spain, 11–15 July 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 1–6.
52. Ying, Z.; Li, G.; Ren, Y.; Wang, R.; Wang, W. A new image contrast enhancement algorithm using exposure fusion framework. In Proceedings of the International Conference on Computer Analysis of Images and Patterns, Ystad, Sweden, 22–24 August 2017; Springer: Berlin, Germany, 2017; pp. 36–46.
53. Wang, R.; Zhang, Q.; Fu, C.W.; Shen, X.; Zheng, W.S.; Jia, J. Underexposed photo enhancement using deep illumination estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 6849–6857.
54. 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](#)]
55. 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, 13–19 June 2020; pp. 2281–2290.
56. Ignatov, A.; Kobyshev, N.; Timofte, R.; Vanhoey, K.; Van Gool, L. 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.
57. Zhou, S.; Li, C.; Loy, C.C. LEDNet: Joint Low-light Enhancement and Deblurring in the Dark. *arXiv* **2022**, arXiv: 2202.03373.