Unsupervised Low-Light Image Enhancement via Virtual Diffraction Information in Frequency Domain

Xupei Zhang *, Guanghao Wang, Hanlin Qin *, Yue Yu *, Xiang Yan, Shanglin Yang

Posted Date: 12 June 2023

doi: 10.20944/preprints202306.0787.v1

Keywords: Low-light Image Enhancement; Unsupervised Learning; Physics-inspired Computer Vision

Article

Unsupervised Low-Light Image Enhancement via Virtual Diffraction Information in Frequency Domain

Xupei Zhang *, Guanghao Wang, Hanlin Qin *, Yue Yu *, Xiang Yan and Shanglin Yang

School of Optoelectronic Engineering, Xidian University, Xi'an 710071, Shaanxi, China;
* Correspondence: zhangxupei@xidian.edu.cn; Tel.:+86-1577-195-7212 (X.Z.),hlqin@mail.xidian.edu.cn; Tel.:+86-1322-707-9356 (H.Q.),yyu@xidian.edu.cn; Tel.:+86-1819-282-7952 (Y.Y.)

Abstract: With the advent of deep learning, significant progress has been made in low-light image enhancement methods. However, deep learning requires enormous paired training data, which is challenging to capture in real-world scenarios.To address this limitation, this paper presents a novel unsupervised low-light image enhancement method, which first introduces the frequency domain features of images in low-light image enhancement tasks. Our work is inspired by imagining a digital image as a spatially varying metaphoric "field of light", then subjecting the influence of physical processes such as diffraction and coherent detection back onto the original image space via a frequency-domain to spatial-domain transformation (inverse Fourier transform). However, the mathematical model created by this physical process still requires complex manual tuning of the parameters for different scene conditions to achieve the best adjustment. Therefore, we proposed a dual-branch convolution network to estimate pixel-wise and high-order spatial interactions for dynamic range adjustment of the frequency feature of the given low-light image. Guided by the frequency feature from the "field of light" and parameter estimation networks, our method enables dynamic enhancement of low-light images. Extensive experiments have shown that our method performs well compared to state-of-the-art unsupervised, and its performance approximates the level of the state-of-the-art supervised methods qualitatively and quantitatively. At the same time, the light network structure design allows the proposed method to have extremely fast inference speed(near 150 FPS on an NVIDIA 3090 Ti GPU for an image of size $600 \times 400 \times 3$). Furthermore, the potential benefits of our method to object detection in the dark are discussed.

Keywords: low-light image enhancement; unsupervised learning; physics-inspired computer vision

1. Introduction

Image capturing in suboptimal lighting conditions is a common occurrence, leading to images with low brightness, poor contrast, and color distortion, which consequently hinder computer vision tasks including object detection and image segmentation. To combat these issues, low-light image enhancement has emerged as an essential research topic in computer vision, particularly for improving the visual fidelity of sub-optimal photos. However, suboptimal lighting conditions necessitate a comprehensive approach rather than simply amplifying brightness to enhance the contrast, as this may inversely impact the overall quality of the image. Therefore, addressing the fundamental causes of low light imaging is crucial to produce high-quality images that meet the needs of various tasks in computer vision and image analysis.

Various traditional methods have been proposed to mitigate further the degradation caused by low-light conditions. These methods are divided into two main categories. Some of the methods depended on the Retinex theory[1,2] and the others based on the Histogram equalization[3,4]. The Retinex-based method involves the decomposition of images into reflection and illumination components. The first component contains information about the scene's inherent attributes like texture, edge details, and color. Meanwhile, the second component contains distribution information on contours and lighting. On the other hand, the main idea behind histogram equalization methods is to increase the dynamic range of the gray values in an image by adjusting its gray distribution.

 doi:10.20944/preprints202306.0787.v1

It is achieved by rearranging the pixels of the image to improve its overall dynamic range. These proposed methods use image-specific curve mapping instead of randomly changing the histogram distribution or relying on inaccurate physical models, resulting in natural enhancement without creating unrealistic artifacts. However, previous methods may still have limitations, particularly concerning their processing of high-noise pictures and their potential to cause insufficient local brightness enhancement and loss of details.

In recent years, deep convolutional neural networks (CNNs) have established the state-of-the-art in low-light image enhancement due to their ability to learn superior feature representation. Advanced techniques have emerged for image enhancement, such as end-to-end learning methods, methods based on learning the components of illumination, and unsupervised and semi-supervised learning methods. However, most CNN-based methods necessitate paired training data, which is challenging to acquire for the same scene with both low-light and normal-light images. To address this, unsupervised deep learning-based methods have been proposed, but they often rely on carefully selected training data and may not generalize well. Furthermore, deep neural networks may pose challenges for practical applications, mainly due to their high memory footprint and long inference time. Thus, the need arises for deep models with low computational cost and fast inference speed for deployment on resource-limited and real-time devices, such as mobile platforms.

Through the brief survey of the model-based and data-driven methods, it is not difficult to find that three significant challenges in low-light image enhancement still exist, which are listed below.

(1) The model-based methods, which aim to build an explicit model to enhance the low-light images, but the suboptimal lighting conditions dramatically increase model complexity model complexity. Therefore, these methods require complex manual tuning of parameters and even idealization of some mathematical processes, making it challenging to achieve dynamic adjustment and even more difficult to achieve optimal enhancement results;
(2) The data-driven methods typically employ a limited size of convolutional kernels to extract the image feature, which have a limited receptive field to obtain the global illumination information for adaptive image enhancement. Consequently, bright areas in the original image may become overexposed after enhancement processing, leading to poor overall visibility. Furthermore, a natural concern for data-driven methods is the necessity to acquire large amounts of high-quality data, which is very costly and difficult, especially when these data have to be acquired under the real-world illumination condition for the same scenarios;
(3) Moreover, although deep neural networks have shown impressive performance in image enhancement and restoration, their massive parameter leads to large memory requirements and long inference time, making them unsuitable for resource-limited and real-time devices. To address these issues, designing deep neural networks with optimized network structures and reduced parameters is crucial for practical engineering and real-time device applications, where a low computational cost and fast inference speed of deep models are highly desired.

Considering the issues above and inspired by previous works [5–7], this paper explores the integration of physical-based reasoning into the data-driven method of low-light enhancement. Therefore, aiming at the above situation, we propose a novel end-to-end neural network named the Unsupervised Low-Light Image Enhancement via Virtual Diffraction in Frequency Domain (ULEFD). The main contributions are summarized below.

(1) Inspired by the previous work [7], we proposed a novel low-light image enhancement method that mapped the physics occurring from the frequency domain into a deep neural network architecture to build a more efficient image enhancement algorithm. The proposed method can balance broad applications and performance of the model-based and data-driven based method,as well as data efficiency and a large requirement of training data;
(2) Considering strong feature consistency in images under varying lighting conditions, this paper designed an unsupervised learning network based on the recursive-based gated convolution

 doi:10.20944/preprints202306.0787.v1

block to obtain the global illumination information from the low-light image. Furthermore, the unsupervised network is independent of paired and unpaired training data. Through this process, the network is able to extract higher-order, consistent illumination features in images, thus providing support for the global adaptive image enhancement task without the large amounts of high-quality data;

(3) In this paper, the superiority of the proposed unsupervised algorithm is verified by comparative experiments with the state-of-the-art unsupervised algorithms based on the different low-light public datasets. Furthermore, The expansion experiment demonstrated that the ULEFD can be accelerated in both physical modeling and network structure levels while still keeping impressive image enhancement performance, which has great potential for deployment on resource-limited devices for real-time image enhancement.

The rest of this work is structured as follows: Section 2 concerns related work, describing current related approaches to low-light image enhancement and the existing problems; In section 3, the proposed image enhancement method ULEFD is described in detail; Section 4 provides the experimental results and discussion. Meanwhile, the expansion experiments for our method and the comparison methods are also provided in Section 4. Finally, conclusions and future work are drawn in Section 5.

2. Related Work

For decades, low-light image enhancement has received significant attention in computer version tasks. As mentioned above, the mainstream methods for low-light image enhancement can be roughly categorized as model-based and data-driven methods. This section briefly reviews these related works and discusses the inspiration from these methods.

2.1. Model-based methods

Low-light image enhancement is a critical area for image processing, with a range of classical and more recent algorithms developed to improve image quality in low-light conditions[8]. Model-based methods include Gamma Correction[9,10], Histogram Equalization[11–15], and Retinex Theory[1,16–18], each with its strengths and weaknesses. Gamma Correction edits the gamma curve of the image to improve contrast by detecting dark and light segments of the image but struggles with complex global parameter selection and local over-exposure[19]. Histogram Equalization stretches the dynamic range of the image by equally distributing pixel values but can lead to artifacts and unexpected local over-exposure as well[5]. However, Histogram Equalization methods are still widely relied on, despite their tendency to suffer from color distortion and other image artifacts. The Retinex Theory decomposes images into reflectance and illumination maps to estimate and enhance illumination in non-uniform lighting conditions. However, these methods can lead to unrealistic or partially over-enhanced images without carefully accounting for noise and other factors[20]. More recent methods abandon these approaches to employ image-specific curve mapping for light enhancement, which enables broader dynamic range adjustment and avoids creating unrealistic artifacts. In addition, several other model-based approaches, including frequency-based[7] and image fusion[21], are also commonly used to enhance images in low-light conditions. These methods expand the research avenues of low-light image enhancement methods from different perspectives. However, these types of methods also suffer from the inability to achieve adaptive adjustment for low-light images.

2.2. Data-Driven methods

Data-driven methods typically rely on either Convolutional Neural Network (CNN)-based or Generative Adversarial Network (GAN)-based approaches. Most CNN-based methods require paired data for supervised training[20,22–25], which can be resource-intensive to obtain. It often involves collecting data through automated light degradation and altering camera settings during

 doi:10.20944/preprints202306.0787.v1

image acquisition or retouching. To improve the weakness, some CNN-based methods, such as LL-Net[26] and MBLLEN[27], generate synthetic data through gamma correction or photosensitivity changes, while datasets like LOL[22] and MIT-Adobe FiveK[28] collect paired low/normal light images. Retinex-based deep models[23,29–31] are also trained on paired data. Frequency-based decomposition-and-enhancement models, such as [32], use real low-light datasets for training. Nonetheless, these methods are constrained by the amount of paired data required and often yield poor generalization capabilities. In contrast, unsupervised GAN-based[33,34] methods like EnlightenGAN[34] and semi-supervised models like [35] learn to enhance images without paired data, although a careful selection of unpaired data is needed. While such methods eliminate paired data's drawbacks, generalization and overfitting are still challenges. Ultimately, data-driven methods are a promising and constantly evolving field, with ongoing research into overcoming these challenges and improving low-light image enhancement. Furthermore, for most of the data-driven methods, a complex and large-scale network is introduced for image enhancement, and the massive number of parameters makes these methods time-consuming. When applied in real-time applications, significant delays may occur. Table 1 summarizes the main properties of the different types of methods.

Table 1. The main properties of model-based methods and data-driven based methods.

Method	Model-Based	Data-Driven
Advantage	Data efficient	Require limited priors
	Physics are universal	Highly performance
	Resource-friendly	Dynamic adjustment
Disadvantage	Require precise modeling	Careful selection data
	Sub-optimal performance	Efficiency depend on structure
	Unable to adaptive adjustment	Highly computational cost

In summary, model-based methods, which aim to build an explicit model to enhance low-light images, possess resource-friendly properties and impressive data efficiency due to their universal underlying physical rules. However, when applied in different scenarios, these methods must converge to a good enough local optimum through carefully designed hand-crafted priors or specific statistical models. In contrast, data-driven methods can improve the ability of model-based methods to understand and analyze data by incorporating a larger number of parameters. This allows for an implicit representation of enhancement modeling, resulting in a high-quality local optimum when the model is adequately trained. However, it is important to note that these methods require large amounts of carefully selected paired or unpaired data, which are often difficult to obtain. Additionally, these implicit models restrict the scope of their application due to the lack of general model-based reasoning and may suffer from overfitting. On the other hand, some data-driven methods, thanks to their larger number of parameters, are able to dynamically adjust to low-light image enhancement tasks. Nevertheless, this also brings higher computational costs.

3. Materials and Methods

Figure 2 illustrates the detailed structure of ULEFD, which comprises two primary modules: the Brightness Adjustment in Frequency Domain(BAFD) component and the Global Enhancement Net(GEN) component. The BAFD component takes the L channel as an input and transforms the L channel from the spatial domain to the frequency domain. Inspired by the [7], the digital image can be

 doi:10.20944/preprints202306.0787.v1

reimaged as a spatially varying metaphoric "field of light."After transferring this "field of light" to the frequency domain, it can provide image brightness adjustment information from physical processes such as diffraction and coherence detection. In addition, to overcome the problem, the original physical brightness adjustment model requires complex manual tuning of parameters in various scenes, and the adjustment effect can only converge to suboptimal results. We used a lightweight network architecture to extract the L channel feature to achieve dynamic adjustment of the turning of paraments.

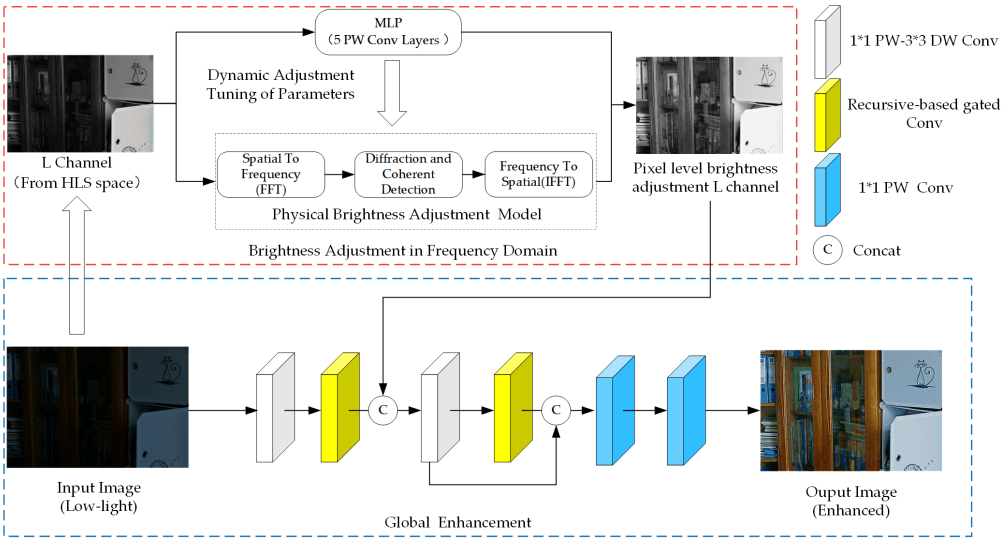

Figure 1. The detailed structure of the proposed method.

The GEN component takes the low-light image and the dynamic brightness adjustment proposal as inputs and enhances the image with some carefully designed loss functions. It consists of different types of convolutional layers, especially the recursive-based gated with a variable receptive field to capture local and global image information and generate high-order spatial information interaction for better performance of the low-light enhancement.

3.1. Brightness Adjustment in Frequency Domain Component

3.1.1. Physical Brightness Adjustment

In [7], the authors demonstrate that introducing the concepts of virtual light field to use the frequency domain information of images as low-light image enhancement has significant effects. Specifically, let the $I(x,y)$ be the original spatial domain digital image.The virtual “field of light” of $I(x,y)$ can be represent as:

$$I(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \widetilde{I}_i(k_x,k_y)e^{+j(k_x,k_y)}dk_xdk_y \tag{1}$$

where the $\widetilde{I}_i(k_x,k_y)$ represents the spatial spectrum of the virtual "field of light". Then, the brightness gain can be obtained by transforming the spatial signal to the frequency domain, and this gain can be represented as a spectral phase:$\phi(k_x,k_y)$, the brightness adjustment can be defined as:

$$\widetilde{I}_o(x,y) = \widetilde{I}_i(k_x,k_y)e^{-i\phi(k_x,k_y)}dk_xdk_y \tag{2}$$

In the end, the brightness gain in the frequency domain needs to be mapped back to the image in the normal spatial domain as:

$$I_o(x,y) = IFFT\{\widetilde{I}_i(k_x,k_y)e^{-i\phi(k_x,k_y)}\} \tag{3}$$

 doi:10.20944/preprints202306.0787.v1

where $IFFT$ refers to the inverse fourier transform, and the $I_o(x, y)$ now contains frequency-dependent brightness gain entirely described by the phase function $\phi(k_x, k_y)$.

As known, digital images are has three bands corresponding to the three fundamental color channels (RGB). However, when performing low-light image enhancement, it is necessary to adjust the image brightness range while preserving the original color saturation information of the image. This requires separating the color information from the luminance information to the greatest extent possible. Therefore, we tried different color space conversion methods to keep the image color saturation information as much as possible and adjust only the image brightness information. As shown in the figure 2,through experiments, we found that brightness adjustment of the image in HLS space[36] has the best effect on preserving the original color saturation information of the image.

Figure 2. Ablation study of the advantage of HLS color space.

3.1.2. Mathematical Modeling

Given our focus on digital images, we transition from a continuous-valued $I(x, y)$ in the spatial domain to a pixelated waveform $I[n, m]$. In the frequency domain, the discrete waveform $I[n, m]$ is expressed as a sum of complex exponential waves with different frequencies:

$$I[n,m] = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \hat{I}[k,l] e^{j2\pi(\frac{kn}{N} + \frac{lm}{N})} \tag{4}$$

where N is the number of pixels in each dimension,j is the imaginary unit, and $\hat{I}[k, l]$ is the discrete Fourier transform (DFT) of $I[n, m]$ defined as:

$$\hat{I}[k,l] = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} I[n,m] e^{-j2\pi(\frac{kn}{N} + \frac{lm}{N})} \tag{5}$$

Similarly, we shift from continuous (kx, ky) to discrete momentum $[kn, km]$.

Therefore,the Gaussian function with zero mean and variance T can be used for the phase function $\phi(k_x, k_y)$ transformation as :

$$\phi[kn, km] = S \cdot \hat{\phi} \tag{6}$$

Resulting in a spectral brightness adjustment operator,

$$H[kn, km] = e^{-i\phi[kn,km]} = e^{-iS\cdot\hat{\phi}} \tag{7}$$

 doi:10.20944/preprints202306.0787.v1

where S is a parameter that maps the loss or gain of spectral brightness adjustment.

Following the spectral intensity adjustment and inverse Fourier transform operation, coherent detection generates the real and imaginary parts of the optical field. The combined processes of diffraction with the low pass spectral phase and coherent detection produce the output of the physical brightness adjustment model:

$$I_o[n,m] = angle\langle IFFT\{e^{-iS\cdot\hat{\phi}} \cdot FT\{I[n,m]\}\}\rangle \tag{8}$$

where FT denotes the Fourier transform operation, and the $angle$ processes the computation of the phase from a complex-valued function of its argument.

In summary, in order to use the interference information obtained in the frequency domain space at different phases as the brightness adjustment gain of the digital input image, we first add a small constant bias term b to the light field corresponding to the input image $I_i[n,m]$ to make the numerical calculation more stable and to achieve the effect of noise reduction. Then, the input image in the spatial domain is transformed to the frequency domain by the FFT and subsequently multiplied with the complex exponential elements, the parameters of which define the frequency-dependent phase. The inverse Fourier transform (IFFT) is then used to return a complex signal in the spatial domain. Mathematically, the inverse tangent operation in phase detection behaves like an activation function. Before calculating the phase, the signal is multiplied by a parameter called the phase activation gain G. The output phase is normalized to match the image formatting convention [0-255]. This output is then injected into the original image as a new L channel in HSL color space (for low light enhancement). Thus, the output of the physical brightness adjustment model can be represented as:

$$Enhance_l = tan^{-1}(G * \frac{ImI_o[n,m]}{ReI_o[n,m]}) \tag{9}$$

where $ImI_o[n,m]$ and $ReI_o[n,m]$ is the imaginary and real component of $I_o[n,m]$.

3.1.3. Dynamic Adjustment Tuning

The established brightness adjustment model contains three adjustable parameters: the mapping parameter S, bias term b, and phase gain parameter G. The parameters mentioned earlier need manual adjustment to enhance low-light images under varied conditions. Inspired by previous work[5], we propose to extract global information from the L channel of the low-light image and use a five-layer multi-layer perception to learn the parameters as mentioned earlier from the sufficient dataset. This processing can be represented as:

$$\{S,b,G\} = MLP(I^l) \tag{10}$$

where I^l represents the L channel of the low-light image I in the HLS color space,and $MLP(\cdot)$ represents the processing of learning these parameter via the five-layer multi-layer perception.

After obtaining the pixel brightness adjustment proposal in the L channel, it will be concatenated with the middle layer of GEN and fed into the GEN for further enhancement. The entire ULEFD is trained end-to-end, which means that all the components are trained jointly to optimize the overall performance of the network.

3.2. Global Enhancement Net

When utilizing traditional convolutional kernels for image feature extraction, the limited perceptual fields make it challenging for the network to comprehensively understand the image. Moreover, the enhanced image is exceptionally vulnerable to noise as there is a lack of information in the low-illumination image.To address these issues, this paper proposes a Global Light Enhancement Net containing three different convolution structures.

 doi:10.20944/preprints202306.0787.v1

As figure 1 shows, firstly, the Point-wise(1×1 kernel size) and Depth-wise(3×3 kernel size) convolution block is used to extract the input low-light image feature. More specifically, the 1×1 point-wise convolution is applied to aggregate pixel-level cross-channel context, then 3×3 depth-wise convolution to encode channel-level spatial context. This convolution structure has been applied in the state-of-the-art image restoration methods[37,38], proving its effectiveness in image noise reduction.

The essential operation in CNN is "convolution," which provides local connectivity and translation equivariance, these features that bring efficiency and versatility to CNNs. However, while enhancing low-light images, the consistency of the original images in terms of color, contrast, and other image information should be ensured. The small size of conventional convolutional kernels limits their field of perception and thus cannot model long-range pixel correlations, making it difficult to retain consistent information about the global image.To address this challenge, this paper introduces a recursive gated depth convolutional neural network[39], which focuses on using the recursive gated convolution to higher-order interaction of image information and long-distance image information modeling. Benefitting from these abilities,the network is able to avoid severe noise distortion and color degradation when enhancing the dark regions on the input low-light images.

3.3. Loss Function

Due to the lack of absolute supervision information to guide the training process, it is tough to recover these two components from low-light images. The only way is to use relative information in loss function designing, which reduces the assumption of the existence of absolute ground-truth data. Previous unsupervised methods have proposed some useful loss functions, such as normalized gradient loss [40], spatial consistency loss[5,6] and perception loss[34]. However, only some achieve impressive results, mainly due to the ineffective use of more specific constraint information in designing these loss functions. Therefore, in this paper, we design each loss function of the algorithm for the image feature information in different components.

3.3.1. Loss for Brightness Adjustment in Frequency Domain component

First, for the component of brightness adjustment in the frequency domain, Low-light degradation causes changes in pixel intensity and color distribution of images. Therefore, we adopt the image color histogram prior to constraining the dynamic brightness adjustment. Specifically, we define an MSE loss inspired by[19,41]. The main idea of this loss function design is that the color histogram prior information contains not only the input low-light image's color distribution information but also the image's structural and semantic information at the higher level, which can be extracted from this color distribution information. The kernel density estimation has been used to keep the loss differentiable:

$$L_{hist} = \frac{1}{N} \sum_{i=1}^{N} \parallel Hist(I_{en}^{i}) - Hist(I_{low}^{i}) \parallel_2^2 \tag{11}$$

where the N means the batch size of the input, the I_{low}^{i} represent the input low-light image and the I_{en}^{i} represent the enhanced image.

In addition, the image maintains its natural and explicit detail content to make the brightness adjustment, and the smooth illumination loss function L_{si} is designed. The main idea is to make the model more focused on image edges and textures by processing the gradient information of the low-light and enhanced images. More specifically, the loss function consists of two different components. The first component is the gradient loss calculation along the x and y directions.

 doi:10.20944/preprints202306.0787.v1

$$\begin{aligned}\mathcal{L}_x &= \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\mathrm{ReLU}(G(R^{low})_{i,j})\exp(-10G(R^{low})_{i,j})\exp(-10G(I)_{i,j})\\ \mathcal{L}_y &= \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\mathrm{ReLU}(G(R^{low})_{i,j})\exp(-10G(R^{low})_{i,j})\exp(-10G(I)_{i,j})\end{aligned} \tag{12}$$

where $G(R^{low})$ represents the normalization of the gradient of the low-illumination image,$G(I)$ represents the normalization of the gradient of the image after enhancement.Moreover,ReLU means the rectified linear unit function.

$$\mathrm{ReLU}(x) = \max(x, 0) \tag{13}$$

The other component of the L_{si} is:

$$\mathcal{L}_{smooth} = (\|G(R^{low}) - G(I))\|_p + \varepsilon)/(CWH) \tag{14}$$

where $\|G(R^{low}) - G(I))\|$ represents the absolute value of the difference between the gradient of the enhanced image and the low-light image.The p means the parametric number (e.g. L1-norm or L2-norm), ε is a very small constant (e.g. $1e-4$) and C is the number of the image channel, H and W is the height and weight of the image.In summary, the total L_{si} loss function is:

$$\mathcal{L}_{si} = \mathcal{L}_x + \mathcal{L}_y + \mathcal{L}_{smooth} \tag{15}$$

3.3.2. Loss for Global Enhancement component

From two aspects of maintaining image color and contrast consistency, two loss functions are applied in this paper for global light enhancement. The first loss function is color constancy loss. The main idea is calculating the mean channel value for both the enhanced image and the input low-light image to obtain the average pixel values of the enhanced image $enhances_{i,j,c}$ and the input low-light image $originals_{i,j,c}$. The processing can be defined as follow:

$$\begin{aligned}enh\_cols_c &= \frac{\sum_{i=1}^{H}\sum_{j=1}^{W} enhances_{i,j,c}}{H \times W}\\ ori\_cols_c &= \frac{\sum_{i=1}^{H}\sum_{j=1}^{W} originals_{i,j,c}}{H \times W}\end{aligned} \tag{16}$$

Then the ratio difference between the three different color channels is calculated as follow.

$$\begin{aligned}rg\_ratio &= \left|\frac{enh\_cols_r}{enh\_cols_g} - \frac{ori\_cols_r}{ori\_cols_g}\right|\\ gb\_ratio &= \left|\frac{enh\_cols_g}{enh\_cols_b} - \frac{ori\_cols_g}{ori\_cols_b}\right|\\ br\_ratio &= \left|\frac{enh\_cols_b}{enh\_cols_r} - \frac{ori\_cols_b}{ori\_cols_r}\right|\end{aligned} \tag{17}$$

The final color consistency loss is obtained by summing the above three ratio differences and taking the mean value of the results:

$$L_{col} = \frac{1}{N}\sum_{i=1}^{N}(rg\_ratio_i + gb\_ratio_i + br\_ratio_i) \tag{18}$$

 doi:10.20944/preprints202306.0787.v1

where c means the picture channel (red, green, blue), the H and W are the height and width of the image, the N is the number of images.

To preserve the contrast consistency, we add a gradient consistency loss. The main idea is to extract the gradients of each channel and calculate the gradient consistency loss by comparing the similarity of the corresponding gradients in the original and enhanced images. The gradient consistency loss can be represented as:

$$L_{grad} = \frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{e\bar{n}h_i^c \cdot o\bar{r}i_i^c}{\|e\bar{n}h_i^c\| \cdot \|o\bar{r}i_i^c\| + 0.00001}\right) + \frac{1}{N}\sum_{i=1}^{N}\cos^{-1}\left(\frac{e\bar{n}h_i^c \cdot o\bar{r}i_i^c}{\|e\bar{n}h_i^c\| \cdot \|o\bar{r}i_i^c\| + 0.00001}\right) \tag{19}$$

where i means the number of images and c means the color channel of the images.

In the end, we use an exposure control loss (L_{exp}) to control the exposure level and avoid under-/over-exposed regions. This loss function quantifies the difference between the average intensity value of a local region and the desired level of well-exposedness (E). The calculation of this loss function consists of the following main steps. First, The enhanced image is fed into the function, which performs an averaging pooling operation and calculates its grayscale value, obtained by averaging the pixel values of the red, green, and blue channels.

$$\text{avg\_intensity} = \frac{1}{r^2}\sum_{i=1}^{r^2}\left(\frac{R_i + G_i + B_i}{3}\right) \tag{20}$$

where r means the window size of pooling operation,R_i, G_i, B_i are represent the color channel of the image. Then calculating the difference between the average grayscale value and the given threshold, and take the absolute value and then average to obtain the exposure control loss as follow.

$$\mathrm{L_{exp}} = \frac{1}{n}\sum_{i=1}^{n}|\text{avg\_intensity}_i - E| \tag{21}$$

where n means the number of window for pooling operation.avg_intensity_i represents the average value of the ith pooling window,and the E is the given threshold.

In summery,the total loss function for the proposed method can be expressed as follow:

$$\mathrm{L} = W_{hist}L_{hist} + W_{si}L_{si} + W_{col}L_{col} + W_{grad}L_{grad} + W_{exp}L_{exp} \tag{22}$$

where the weights $W_{hist}, W_{si}, W_{col}, W_{grad}, W_{exp}$ are used for balancing the scales of different losses.

4. Experiment and Results

In this section, we present the implementation details of our proposed low-light image enhancement method. Afterward, we perform both qualitative and quantitative comparisons with state-of-the-art supervised and unsupervised methods, utilizing traditional metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM)[42], and Natural Image Quality Evaluator (NIQE)[43]. In addition, we conduct ablation studies to demonstrate the effectiveness of each component or loss in the proposed method. Finally, we investigate the performance of our method to improve the efficiency of downstream tasks, such as face detection in the dark.

4.1. Implementation Details

The framework is implemented with PyTorch on an NVIDIA 3090 Ti GPU with 24GB memory. The batch size used for training is 64. We use the Adam optimizer to train the network with an initial learning rate of $1e^{-4}$ and decate rate of 0.5 every 50 epochs. We mainly use two datasets for training and comparisons: the LOL dataset[22] and VE-LOL dataset[44].

 doi:10.20944/preprints202306.0787.v1

4.2. Quantitative Evaluation

In this section, we compared our method with several state-of-the-art low-light image enhancement methods. These methods include: one conventional method (LIME[17]), two supervised methods(KinD++[24], Restormer[37]) and three unsupervised methods(Zero-DCE++[5], EnlightGAN[34], LE-GAN[45]).To demonstrate the robustness of our proposed method, we give more experiments on cross-dataset. We have fine-tuned all the above methods on the train sets of LOL and VE-LOL datasets and then evaluated them on their test sets. From Table 2, our method achieves significantly better results among all unsupervised methods, and its performance approximates the level of the state-of-the-art supervised methods. It is obvious that the proposed ULEFD can achieve better PSNR than other unsupervised methods and some supervised methods, whether trained on the LOL or VE-LOL dataset. Regarding SSIM, the proposed method achieved results close to the supervised methods KinD++[24] and Restormer[37], which does not require any reference images for training. However, the proposed method has fewer parameters (only 70K parameters) and costs less running time during testing.

Table 2. Quantitative comparison results on (LOL[22] & VE-LOL[44]) datasets.Red and blue indicate the best and the second-best results, respectively.

Learning	Method	LOL			VE-LOL			Efficiency	
		PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	Params(M)↓	test time(s)↓
Conventional	LIME[17]	16.76	0.56	10.61	14.77	0.53	10.85	-	0.491(on CPU)
Supervised	KinD++[24]	21.30	0.82	11.02	20.87	0.80	11.60	8.28	0.829
	Restormer[37]	23.17	0.84	10.14	22.49	0.82	10.53	8.19	0.821
Unsupervised	Zero-DCE++[5]	14.86	0.57	10.95	16.93	0.68	10.81	0.01	0.0012
	EnlightGAN[34]	16.21	0.59	14.74	17.48	0.65	14.42	8.63	0.871
	LE-GAN[45]	21.38	0.82	11.32	21.50	0.82	10.71	8.92	0.907
	our(Training on LOL)	21.97	0.83	10.23	21.63	0.83	10.21	0.07	0.008
	our(Training on VE-LOL)	21.44	0.82	10.19	22.12	0.84	10.13	0.07	0.008

To further demonstrate the generalization ability of the proposed method, we have tested the proposed method on some real-world low-light image sets, including DICM[14](64 images), LIME[17](10 images), VV[1](24 images), LCDP[46], SCIE[47](select 100 low-light images from the datasets).In the expanding experiments, we use unpaired public datasets and the NIQE metric to compare the proposed method quantitatively with state-of-the-art methods that assess natural image restoration without requiring ground truth. Table 3 contains the NIQE scores for five different public datasets that were previously used in relevant studies. In summary, these experimental results show the effectiveness of our proposed method.

[1]https://sites.google.com/site/vonikakis/datasets

Table 3. NIQE scores on low-light image sets(DICM[14], LIME[17], VV[1], LCDP[46], SCIE[47]). The best result is in red whereas the second best results are in blue, respectively. Smaller NIQE scores indicate a better quality of perceptual tendency.

Learning	**Method**	DICM[14]	LIME[17]	VV[1]	LCDP[46]	SCIE[47]	Avg
Conventional	LIME[17]	11.823	10.612	11.672	9.456	10.818	10.876
Supervised	KinD++[24]	15.043	10.911	11.449	9.461	11.451	11.663
	Restormer[37]	14.012	10.290	11.128	9.352	10.787	11.114
Unsupervised	Zero-DCE++[5]	10.995	10.932	10.645	10.217	10.56	10.70
	EnlightenGAN[34]	15.201	11.335	11.298	9.251	10.546	11.526
	LE-GAN[45]	11.928	10.69	10.41	10.364	10.588	10.796
	Our	10.037	10.084	10.504	9.336	10.245	10.041

4.3. Qualitative Evaluation

Figure 3 shows some representative results for visual comparison from LOL dataset. We have zoomed in the details inside the red and green bounding boxes to further investigate the differences between these comparison methods.The enhanced results show that the conventional method LIME[17]

 doi:10.20944/preprints202306.0787.v1

enhances the images by directly estimating the illumination map but has some external noises. For unsupervised methods, Zero-DCE++[5] produces under-enhanced and noisy results, respectively. Meanwhile, the KIND++[24] has apparent noise and weak illumination.EnlightenGAN[34] suffers from under-enhanced and over-smoothing. The LE-GAN[45] performs better than the EnlightenGAN but is still under-enhanced in some local details. Benefiting from the introduction of the normal illumination reference image, the image enhancement effect of the Restormer[37] is closest to the ground truth. In contrast, Figure 3 shows that our method can well preserve the structural and textural image details without reference images to guide the network. It demonstrates that our proposed method achieves more satisfactory visualization results than the unsupervised learning methods for comparison, especially in the exposure level, structure description, and color saturation.

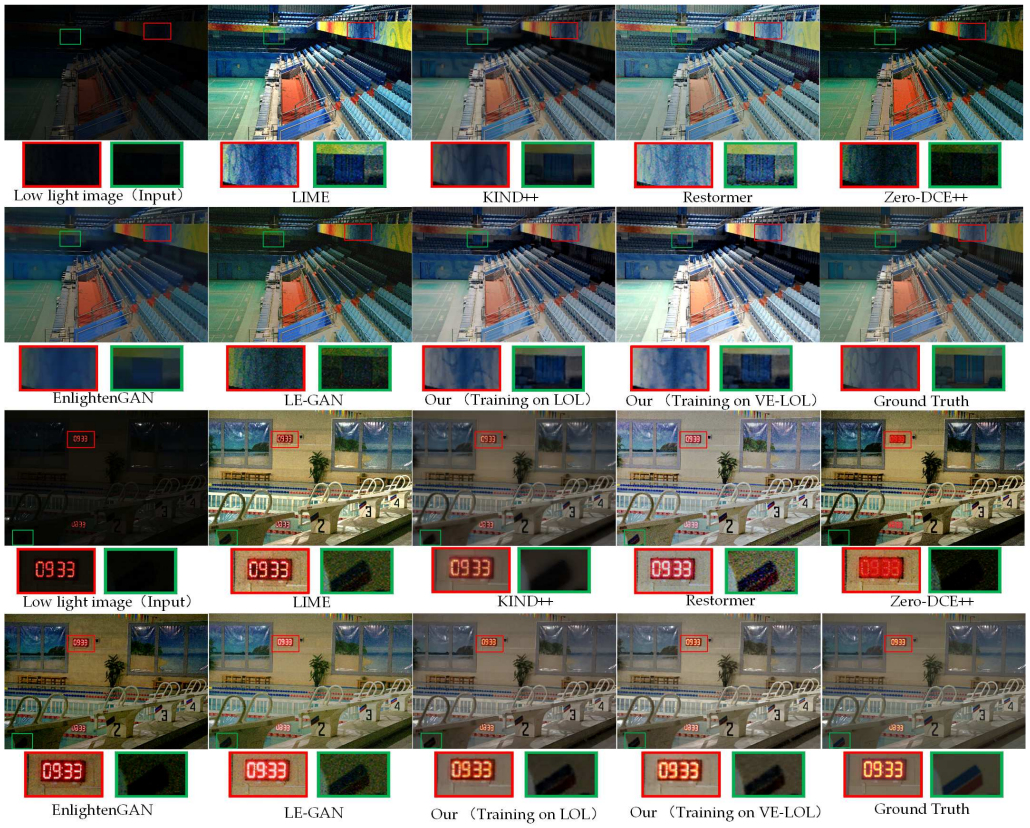

Figure 3. Qualitative results on LOL test dataset.

Figure 4 shows some representative results for visual comparison from the VE-LOL dataset. This dataset further expands the scenario based on the LOL dataset. The enhanced results show that the LIME[17] has severe contrast and noise issues. For unsupervised methods, the results of Zero-DCE++[5] also suffer from extreme contrast and noise issue. KIND++[24] has weak illumination.EnlightenGAN[34] still suffers from under-enhanced and over-smoothing. Regarding the LE-GAN[45], the global enhancement effect is better than the above methods, but there are some issues of color distortion in a few details. In terms of global and local effects of image enhancement, the proposed method in this paper, especially the model trained on the VE-LOL training set, is able to obtain almost the same enhancement results as the Restormer[37], which is the supervised learning method, achieving visual quality close to the ground truth.

 doi:10.20944/preprints202306.0787.v1

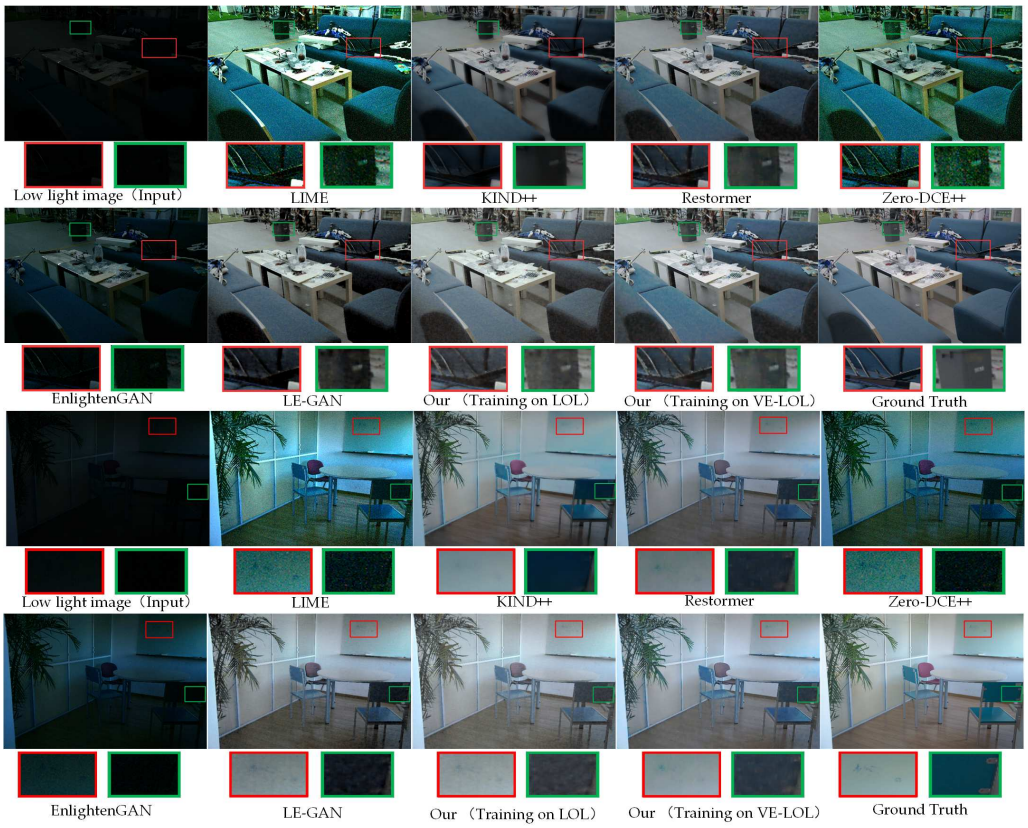

Figure 4. Qualitative results on VE-LOL test dataset.

Figure 5 shows the image enhancement effect of the algorithm in this paper and other comparison algorithms in real low-light scenarios, respectively.Zero-DCE++[5] fail to suppress noise when the background of the scenarios is extremely dark in the DICM[14] and LIME[17] datasets. Meanwhile, EnlightenGAN[34] provided limited image enhancement in the above scenarios.KIND++[24] suffer from blurring artifacts in LIME[17] dataset. As for LCDP[46] datasets, Zero-DCE++[5] and LE-GAN[45] easily lead to over-exposure artifacts and blurriness, which make the results distorted and glaring with information loss. LIME[17] retains the contrast information of images in all of the datasets relatively well, but the overall enhancement effect is weak. In contrast, our proposed method in all datasets tends to generate the same performance as the state-of-the-art supervised method Restormer[37], with proper color contrast, sufficient detailed information, and acceptable and controllable noise.

 doi:10.20944/preprints202306.0787.v1

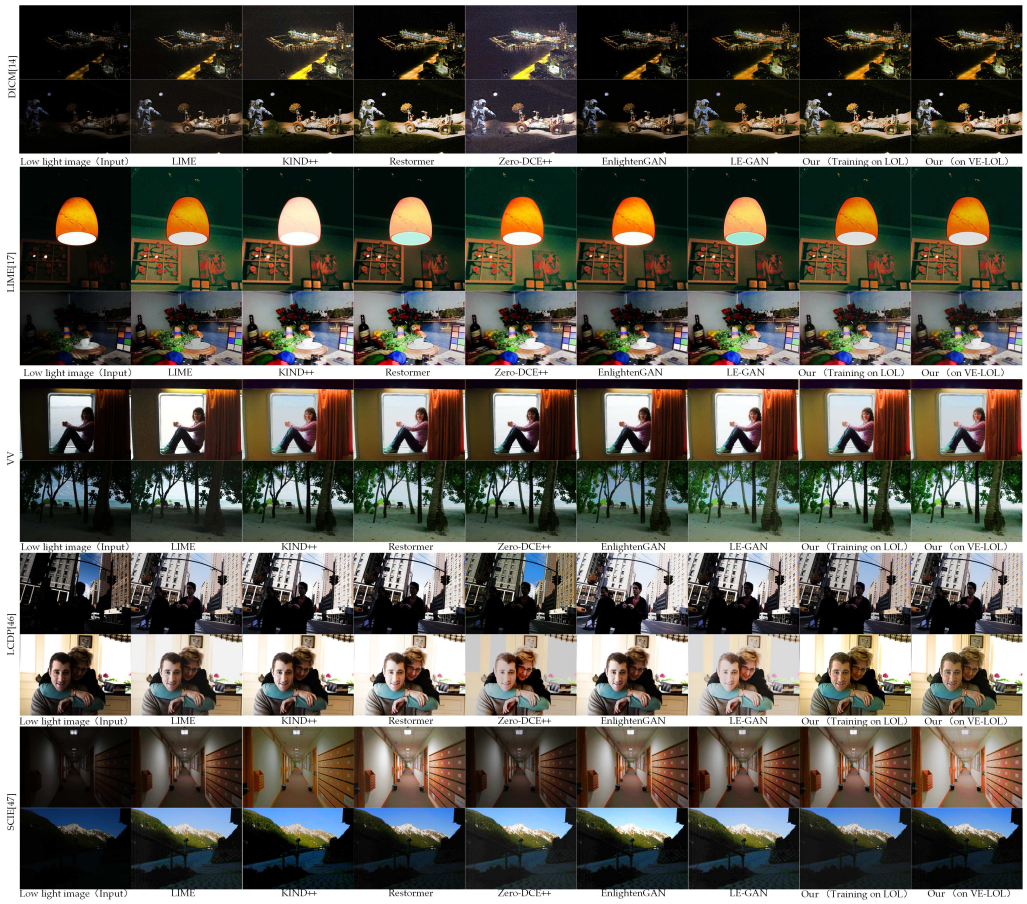

Figure 5. Qualitative results on DICM[14], LIME[17], VV[1], LCDP[46] and SCIE[47] datasets, respectively.

4.4. Ablation Study

4.4.1. Contribution of BAFD component:

In this ablation study, the network only has the GEN component, and the three associated loss function L_{col},L_{grad}, and L_{exp} are considered as the baseline model. The effects of adding the BAFD component and losses proposed in this paper were compared and studied. The results are presented in Table 4.

Table 4. The influence of BAFD component and loss functions based on relative information. During training. Relative losses represents $L_{col} + L_{grad} + L_{exp}$

Loss functions			BAFD	LOL		VE-LOL	
L_{hist}	L_{si}	Relative losses	component	PSNR	SSIM	PSNR	SSIM
		✓		17.52	0.80	18.87	0.73
✓		✓	✓	19.05	0.81	19.42	0.82
	✓	✓	✓	20.39	0.82	21.55	0.83
✓	✓	✓	✓	**21.44**	**0.82**	**22.12**	**0.84**

From Table 4, it can be observed that when we add the other losses proposed in this paper or the BAFD component to the baseline model, both PSNR and SSIM show improvement.This proves the effectiveness of the BAFD component and the loss functions designed with relative information. The BAFD component can adjust the global brightness information and integrate it into the enhancement process with few parameters, which can effectively improve the PSNR by 2.87 dB and the SSIM value by 0.02 (PSNR: 17.52 $\rightarrow$ 20.39, SSIM: 0.74 $\rightarrow$ 0.82).

 doi:10.20944/preprints202306.0787.v1

4.4.2. Contribution of Each Loss:

In this ablation study, We present the results of ULEFD trained by various combinations of losses. As shown in Table 5, the performance of the proposed ULEFD steadily increases with the addition of five loss functions, and the effectiveness of our hybrid loss function is proved. As shown in Figure 6, The result without the BAFD component has limited brightness adjustment than the full result. The result of smooth illumination loss L_{si} has a relatively lower color contrast than the full result. Severe color casts emerge when the histogram prior loss L_{hist} is discarded.

Table 5. The influence of different training losses.

Loss functions					LOL		VE-LOL	
L_{hist}	L_{si}	L_{col}	L_{grad}	L_{exp}	PSNR	SSIM	PSNR	SSIM
✓					12.62	0.54	14.26	0.57
✓	✓				17.88	0.68	18.49	0.70
✓	✓	✓			18.24	0.70	18.86	0.71
✓	✓	✓	✓		20.72	0.77	21.60	0.79
✓	✓	✓	✓	✓	**21.44**	**0.82**	**22.12**	**0.84**

Meanwhile, it hampers the correlations between neighboring regions leading to apparent artifacts. Removing the color constancy loss L_{col} fails to recover the color contrast of the image. Removing the gradient consistency loss L_{grad} hampers the correlations between neighboring regions leading to apparent artifacts. Finally, Removing the exposure control loss L_{exp} fails to brighten the image compared with the full result. Such results demonstrate that the BAFD component and each loss used in the proposed method play a significant role in achieving the final visually pleasing results.

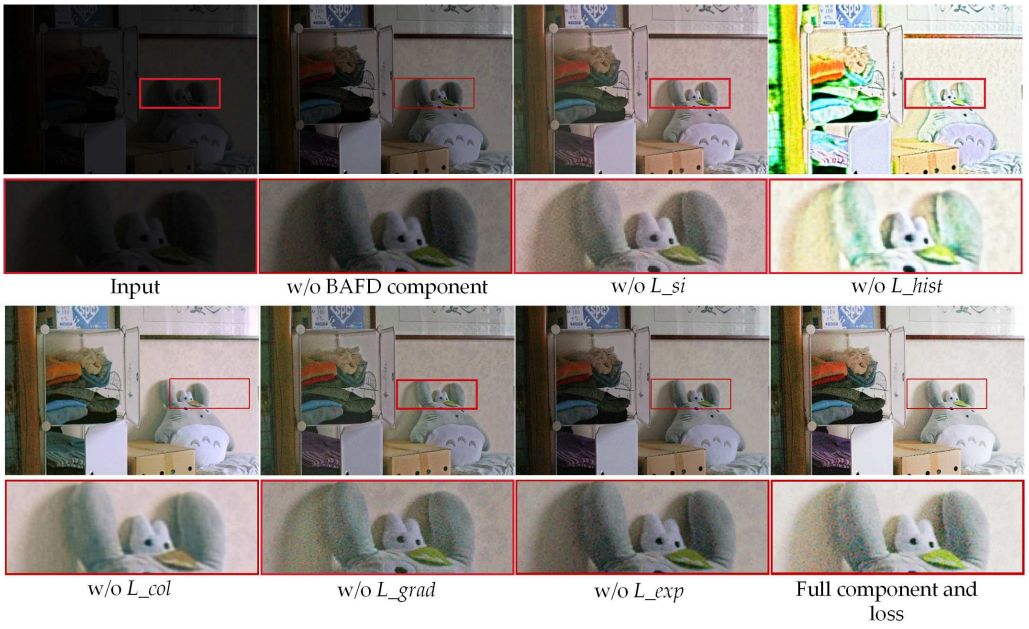

Figure 6. Ablation study of the contribution of BAFD component and each loss (histogram prior loss L_{hist}, smooth illumination loss L_{si}, color constancy loss L_{col},gradient consistency loss L_{grad} exposure control loss L_{exp}).Red boxes indicate the obvious differences and amplified details.

4.5. Pedestrian Detection in the dark

In this section, we aim to evaluate the effectiveness of low-light image enhancement methods for the pedestrian detection task in low-light conditions. We utilized the DARK FACE dataset[48], which consists of 10,000 images captured in low-light conditions. Since the label of the test set is not accessible to the public, we opt to evaluate the proposed method on the training and validation sets

 doi:10.20944/preprints202306.0787.v1

comprising 6,000 images. We adopted the public deep face detector, Dual Shot Face Detector(DSFD)[49], which pre-trained on the WIDER FACE dataset[50], to serve as our baseline model. The results of various low-light image enhancement methods were fed to the DSFD[49] for analysis. We utilized the evaluation tool5 from the DARK FACE dataset[48] to compare the average precision (AP) at various IoU thresholds, including 0.5, 0.7, and 0.9. Table 6 shows the detailed AP results of our evaluation.

Table 6. The average precision (AP) for face detection in the dark under different IoU thresholds (0.5, 0.7, 0,9). The best result is in red whereas the second best one is in blue under each case.

Method	IoU thresholds		
	0.5	0.7	0.9
low-light image	0.231278	0.007296	0.000002
LIME[17]	0.293970	0.013417	0.000007
KinD++[24]	0.243714	0.008616	0.000003
Restormer[37]	0.304128	0.017581	0.000007
Zero-DCE++[5]	0.289232	0.014772	0.000006
EnlightenGAN[34]	0.276574	0.015545	0.000003
LE-GAN[45]	0.294977	0.017107	0.000005
Ours	0.303135	0.017204	0.000009

Based on the results presented in Table 6, it is evident that all the methods' AP scores decrease as the IoU thresholds increase. At an IoU threshold of 0.9, all the approaches perform exceptionally poorly. However, under IoU thresholds of 0.5 and 0.7, the proposed method achieves similar AP scores that are only slightly lower than Restormer[37] superior performance. Moreover, our method achieves balanced subject enhancement performance, application performance, and computational cost without using paired training data. The proposed method effectively lights up facial features in dark areas while preserving features in well-light areas, ultimately improving pedestrian detection in low-light conditions. Figure 7 shows examples of object detection using the Dual Shot Face Detector(DSFD) on low-light images and enhanced images with the proposed method.

Figure 7. Impact of VEViD preprocessing on pedestrian detection in the dark.

 doi:10.20944/preprints202306.0787.v1

5. Discussion

1. Deep-learning-based methods have recently attracted significant attention in the image processing field. Due to the powerful feature representation ability of the data, data-driven methods can learn more general visual features. This property means these methods can be used to relieve some challenges for image enhancement, such as poor illumination conditions. Our research aims to combine the physical brightness adjustment model based on frequency information with a data-driven-based low-light image enhancement method to improve the performance of the dynamic enhancement for low-light images. Moreover, the proposed method is based on a lightweight network design, offering it the advantages of a flexible generalization capability and real-time inference speed. The quantitative results in Table 2 and Table 3 show that the data-driven methods have better image enhancement results on all the test sets than the conventional method when the training data is sufficient. It is due to the fact that the data-driven approach relies on the powerful feature extraction capability of the deep learning network to adjust the brightness of each pixel in the image dynamically. As for data-driven methods, supervised learning usually has better image enhancement results because it can rely on normally exposed images to guide network learning. However, collecting pairs of images in natural environments is very time-consuming. The data dependence of supervised learning also causes a lack of generalization ability of the model. Specifically, the model degrades in scenarios with significant differences from the training data. In contrast, unsupervised learning reduces the reliance on paired data and performs better generalization. In particular, the proposed method in this paper maintains better robustness on different test data. It is able to achieve image enhancement results approximating the state-of-the-art supervised learning.

2. Through ablation experiments, this paper analyzes the reasons for the performance improvement of the algorithm from two aspects. First, the ablation experiments demonstrate that this paper uses the two-branch network structure, and the one-way network introduces the channel characterizing the image luminance with the frequency domain feature model under the assumption of the virtual light field, which can effectively achieve the luminance adjustment. Moreover, a lightweight parameter estimation network can achieve dynamic brightness adjustment. Meanwhile, the other network relies on acquiring global image information to preserve the original image structure, color contrast, and other critical information while enhancing the image so that the enhanced image noise can be better suppressed. On the other hand, the contribution of the loss function of constrained unsupervised learning is analyzed in this paper through ablation experiments. Through the structure of the ablation experiment, it is easy to find that for the brightness adjustment branch, the histogram prior information loss function used in this paper can effectively preserve the original distribution of image information while brightness adjustment, thus making it possible to adjust the brightness without losing the original image semantic structure features. On the other hand, the illumination smoothing loss function allows the network to reduce the impact of noise on the overall image enhancement results during the luminance adjustment learning. For the global enhancement branch, this paper constrains the network to retain the high-level image feature information from two aspects: color gradient consistency and image gradient change consistency, so that the enhanced images achieve significant improvement in both the quantitative and qualitative evaluation(in Table 2,3 and Figure 3,4). Meanwhile, the exposure consistency loss further enhances the intuitive image enhancement effect.

3. To analyze the potential of the algorithms in this paper for real-time applications, the paper first compares the parametric quantities and inference implementations of the various algorithms in Table 2. It can be seen that the number of parameters of the proposed method in this paper is better than most of the comparison methods, and the inference speed is only slightly slower than Zero-DCE++[5], which is significantly light-weight and fast for practical applications.

 doi:10.20944/preprints202306.0787.v1

6. Conclusions

In this work, we propose an unsupervised dual-branch network for low-light image enhancement. One network branch uses the frequency domain information of low-light images to achieve dynamic brightness adjustment of images. At the same time, the other focuses on the global image information to dynamically adjust the overall brightness of images while preserving the high-level structural features of low-light images themselves, guiding the network to suppress noise effectively, color contrast differences, and other problems that exist when enhancing low-light images while enhancing images. Moreover, the loss functions designed in this paper can effectively guide the network to make dynamic adjustments while preserving the structural information of low-illumination images. It further enhances the low-light image enhancement effect and can support the performance improvement of downstream tasks. Finally, the lightweight network structure design reduces the number of network parameters and computational complexity. It improves the inference speed of this paper, which gives the proposed method the potential to be used in computing platforms with limited computing power.

In the future, we plan to integrate semantic information into image sequence enhancement and design a more lightweight network architecture. By combining more prior constraints and reducing the computational cost, further accuracy gains in downstream tasks and more practical applications are achievable.

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

Funding: National Natural Science Foundation of China (62174128); Ningbo Natural Science Foundation (2022J185); Xian City Science and Technology Plan Project (No. 21JBGSZ-OCY9-0004, 22JBGS-QCY4-0006)

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors wish to thank the editors and the reviewers for their valuable suggestions.

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

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Networks
ULEFD	Unsupervised Low-light Image Enhancement via Virtual Diffraction in Frequency Domain
BAFD	Bright Adjustment in Frequency Domain
GEN	Global Enhancemnet Net
FT	Fourier Transform
FFT	Fast Fourier Transform
IFFT	Inverse Fourier Transform
MLP	Multi-layer Perception
MSE	Mean-Square Error

References

1. Wang, S.; Zheng, J.; Hu, H.M.; Li, B. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing* **2013**, *22*, 3538–3548.
2. Wang, L.; Xiao, L.; Liu, H.; Wei, Z. Variational Bayesian method for retinex. *IEEE Transactions on Image Processing* **2014**, *23*, 3381–3396.
3. Pisano, E.D.; Zong, S.; Hemminger, B.M.; DeLuca, M.; Johnston, R.E.; Muller, K.; Braeuning, M.P.; Pizer, S.M. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital imaging* **1998**, *11*, 193–200.

4. Pizer, S.M. Contrast-limited adaptive histogram equalization: Speed and effectiveness stephen m. pizer, r. eugene johnston, james p. ericksen, bonnie c. yankaskas, keith e. muller medical image display research group. In Proceedings of the Proceedings of the first conference on visualization in biomedical computing, Atlanta, Georgia, 1990, Vol. 337, p. 1.
5. Li, C.; Guo, C.; Loy, C.C. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2021**, *44*, 4225–4238.
6. Quan, Y.; Fu, D.; Chang, Y.; Wang, C. 3D Convolutional Neural Network for Low-Light Image Sequence Enhancement in SLAM. *Remote Sensing* **2022**, *14*, 3985.
7. Jalali, B.; MacPhee, C. VEViD: Vision Enhancement via Virtual diffraction and coherent Detection. *eLight* **2022**, *2*, 1–16.
8. Li, C.; Guo, C.; Han, L.; Jiang, J.; Cheng, M.M.; Gu, J.; Loy, C.C. Low-light image and video enhancement using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence* **2021**, *44*, 9396–9416.
9. Farid, H. Blind inverse gamma correction. *IEEE transactions on image processing* **2001**, *10*, 1428–1433.
10. Lee, Y.; Zhang, S.; Li, M.; He, X. Blind inverse gamma correction with maximized differential entropy. *Signal Processing* **2022**, *193*, 108427.
11. Coltuc, D.; Bolon, P.; Chassery, J.M. Exact histogram specification. *IEEE Transactions on Image processing* **2006**, *15*, 1143–1152.
12. Ibrahim, H.; Kong, N.S.P. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics* **2007**, *53*, 1752–1758.
13. Stark, J.A. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Transactions on image processing* **2000**, *9*, 889–896.
14. Lee, C.; Lee, C.; Kim, C.S. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE transactions on image processing* **2013**, *22*, 5372–5384.
15. Singh, K.; Kapoor, R. Image enhancement using exposure based sub image histogram equalization. *Pattern Recognition Letters* **2014**, *36*, 10–14.
16. 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 Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2782–2790.
17. Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing* **2016**, *26*, 982–993.
18. Li, M.; Liu, J.; Yang, W.; Sun, X.; Guo, Z. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing* **2018**, *27*, 2828–2841.
19. Zhang, F.; Shao, Y.; Sun, Y.; Zhu, K.; Gao, C.; Sang, N. Unsupervised low-light image enhancement via histogram equalization prior. *arXiv preprint arXiv:2112.01766* **2021**.
20. Fan, S.; Liang, W.; Ding, D.; Yu, H. LACN: A lightweight attention-guided ConvNeXt network for low-light image enhancement. *Engineering Applications of Artificial Intelligence* **2023**, *117*, 105632.
21. Ying, Z.; Li, G.; Ren, Y.; Wang, R.; Wang, W. A new image contrast enhancement algorithm using exposure fusion framework. In Proceedings of the Computer Analysis of Images and Patterns: 17th International Conference, CAIP 2017, Ystad, Sweden, August 22-24, 2017, Proceedings, Part II 17. Springer, 2017, pp. 36–46.
22. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560* **2018**.
23. Zhang, Y.; Zhang, J.; Guo, X. Kindling the darkness: A practical low-light image enhancer. In Proceedings of the Proceedings of the 27th ACM international conference on multimedia, 2019, pp. 1632–1640.
24. Zhang, Y.; Guo, X.; Ma, J.; Liu, W.; Zhang, J. Beyond brightening low-light images. *International Journal of Computer Vision* **2021**, *129*, 1013–1037.
25. Jiang, N.; Lin, J.; Zhang, T.; Zheng, H.; Zhao, T. Low-Light Image Enhancement via Stage-Transformer-Guided Network. *IEEE Transactions on Circuits and Systems for Video Technology* **2023**.
26. Lore, K.G.; Akintayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition* **2017**, *61*, 650–662.
27. Lv, F.; Lu, F.; Wu, J.; Lim, C. MBLLEN: Low-Light Image/Video Enhancement Using CNNs. In Proceedings of the BMVC, 2018, Vol. 220, p. 4.

 doi:10.20944/preprints202306.0787.v1

28. Bychkovsky, V.; Paris, S.; Chan, E.; Durand, F. Learning photographic global tonal adjustment with a database of input/output image pairs. In Proceedings of the CVPR 2011. IEEE, 2011, pp. 97–104.
29. 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 Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 6849–6857.
30. Ren, W.; Liu, S.; Ma, L.; Xu, Q.; Xu, X.; Cao, X.; Du, J.; Yang, M.H. Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing* **2019**, *28*, 4364–4375.
31. 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 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5901–5910.
32. Xu, K.; Yang, X.; Yin, B.; Lau, R.W. Learning to restore low-light images via decomposition-and-enhancement. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 2281–2290.
33. Chen, Y.S.; Wang, Y.C.; Kao, M.H.; Chuang, Y.Y. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6306–6314.
34. 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 transactions on image processing* **2021**, *30*, 2340–2349.
35. Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; Liu, J. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 3063–3072.
36. Saravanan, G.; Yamuna, G.; Nandhini, S. Real time implementation of RGB to HSV/HSI/HSL and its reverse color space models. In Proceedings of the 2016 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2016, pp. 0462–0466.
37. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5728–5739.
38. Chen, L.; Chu, X.; Zhang, X.; Sun, J. Simple baselines for image restoration. In Proceedings of the Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII. Springer, 2022, pp. 17–33.
39. Rao, Y.; Zhao, W.; Tang, Y.; Zhou, J.; Lim, S.N.; Lu, J. Hornet: Efficient high-order spatial interactions with recursive gated convolutions. *Advances in Neural Information Processing Systems* **2022**, *35*, 10353–10366.
40. Zhang, Y.; Di, X.; Zhang, B.; Li, Q.; Yan, S.; Wang, C. Self-supervised low light image enhancement and denoising. *arXiv preprint arXiv:2103.00832* **2021**.
41. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer, 2016, pp. 694–711.
42. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **2004**, *13*, 600–612.
43. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters* **2012**, *20*, 209–212.
44. Liu, J.; Xu, D.; Yang, W.; Fan, M.; Huang, H. Benchmarking low-light image enhancement and beyond. *International Journal of Computer Vision* **2021**, *129*, 1153–1184.
45. Fu, Y.; Hong, Y.; Chen, L.; You, S. LE-GAN: Unsupervised low-light image enhancement network using attention module and identity invariant loss. *Knowledge-Based Systems* **2022**, *240*, 108010.
46. Haoyuan Wang, K.X.; Lau, R.W. Local Color Distributions Prior for Image Enhancement. In Proceedings of the Proceedings of the European Conference on Computer Vision (ECCV), 2022.
47. Cai, J.; Gu, S.; Zhang, L. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing* **2018**, *27*, 2049–2062.
48. Yuan, Y.; Yang, W.; Ren, W.; Liu, J.; Scheirer, W.J.; Wang, Z. UG $^{2+}$ Track 2: A Collective Benchmark Effort for Evaluating and Advancing Image Understanding in Poor Visibility Environments. *arXiv preprint arXiv:1904.04474* **2019**.

 doi:10.20944/preprints202306.0787.v1

49. Li, J.; Wang, Y.; Wang, C.; Tai, Y.; Qian, J.; Yang, J.; Wang, C.; Li, J.; Huang, F. DSFD: dual shot face detector. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5060–5069.
50. Yang, S.; Luo, P.; Loy, C.C.; Tang, X. Wider face: A face detection benchmark. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 5525–5533.