Low-Light Image Enhancement Using the Cell Vibration Model

Xiaozhou Lei, Zixiang Fei, Wenju Zhou, Huiyu Zhou, and Minrui Fei

Abstract—Low light very likely leads to the degradation of an image’s quality and even causes visual task failures. Existing image enhancement technologies are prone to overenhancement, color distortion or time consumption, and their adaptability is fairly limited. Therefore, we propose a new single low-light image lightness enhancement method. First, an energy model is presented based on the analysis of membrane vibrations induced by photon stimulations. Then, based on the unique mathematical properties of the energy model and combined with the gamma correction model, a new global lightness enhancement model is proposed. Furthermore, a special relationship between image lightness and gamma intensity is found. Finally, a local fusion strategy, including segmentation, filtering and fusion, is proposed to optimize the local details of the global lightness enhancement images. Experimental results show that the proposed algorithm is superior to nine state-of-the-art methods in avoiding color distortion, restoring the textures of dark areas, reproducing natural colors and reducing time cost.

Index Terms—Low light, image enhancement, cell vibration model, guided filtering, image fusion.

I. INTRODUCTION

WITH increasing demands for video recordings created from night scenes, where images are acquired without sufficient exposure, people must deal with problems such as low lightness, low contrast and increasing image noise. Image content may be missing or obscured due to darkness. Therefore, suitable image enhancement technologies are being developed to reveal the image information hidden in the dark [1], [9].

In the field of low light image enhancement, one of the classic and extensively utilized methods is histogram-based techniques. Histogram equalization (HE) [2] is a fast yet simple method that adjusts the corresponding distributions of individual RGB channels in a color image. Problems such as overenhancement, color distortion, and evident noise in dark areas cannot be avoided due to a lack of color correlation constraints across channels. To cope with these shortcomings, several HE-based methods have been proposed, such as CLAHE [3], CVC [4], and MLHE [5], [6]. Celik and Tjahjadi [4] reported that the input image contrast can be enhanced by constructing a two-dimensional histogram based on the neighborhood relationships of image pixels. The disadvantage of the method is that color restoration still needs to be improved [7], [8], [29]. Caselles et al. [5] proposed a local contrast enhancement algorithm for shape preservation. The disadvantage is that the image noise is also amplified [6].

Other widely influential methods include model-based approaches, such as the Retinex model [9]–[19], tone mapping method [21]–[26], physical lighting model [28]–[32], and frequency-domain methods [33]–[35]. For the retinex model, the LIME algorithm proposed by Guo [9] has attracted much attention. By building a smooth structure perception model, the illumination consistency can be improved, and finally, a well-structured illumination map is obtained. Gu et al. [16] proposed a fractional-order variational model based on retinex to obtain appropriate illumination estimation results. Hao et al. [19] proposed a semidecoupled Gaussian total variation model to estimate the illumination and reflectance layers. Singh et al. [20] proposed a texturewise adaptive gamma correction method and a quintile-based reflectance calculation method to enhance textures and improve unbalanced/nonuniform illumination. The advantage of this method is that texture enhancements are implemented in a non-iterative way. However, the calculation is time-consuming. Notably, retinex-based methods often use gamma transformations to readjust illumination estimation results.

For tone mapping methods, Ahn and Keum [23] proposed enhancing dark images using a two-step method that included global mapping and local color adjustment. This method can retrieve image details and has less computational cost. Srinivas et al. [26] proposed an adaptive sigmoid transformation function to adjust lightness, and used a Laplace filter to enhance the edge strength. Li et al. [27] presented an algorithm that adaptively adjusts the global contrast of gray images using the bilateral gamma adjustment function and particle swarm optimization. The advantage is its reduction of the influence of uneven illumination on the image quality but it is limited to considering only gray images rather than color images. In addition, because the final enhanced image is obtained by iterative calculations, it is
very time-consuming. For physical lighting models, Jiang and Yao [28] constructed a foggy image by reversing the pixels of the dark image and then used the dark channel algorithm to remove the noise to improve the contrast of the dark image. Based on the illumination-reflection model and multiscale theory, Wang and Chen [29] presented a color correction method using a nonlinear transformation function.

For frequency-domain methods, Zhang et al. [33] enhanced the local contrast information of uneven-illuminated images by combining the homomorphic filtering method with discrete cosine transformation (DCT). The blocking effect of the homomorphic filtering method is reduced in the spatial domain by using gradient information and DCT features. Kawasaki et al. [34] used the LL component of wavelet expansion to estimate the illumination component and used 2-level wavelet decomposition and reconstruction methods to improve the accuracy of the illumination components. This method improves the performance of the traditional MSR methods, but it is time consuming. AAMIR et al. [35] proposed a low light level enhancement method using wavelet transform. A dual-tree complex wavelet is used to decompose V-channel images to obtain high- and low-frequency components. The low-frequency component is enhanced to improve the brightness of the image, and the noise of the high-frequency component is suppressed to strengthen the texture features.

Deep learning technology offers potential solutions to the field of low-light image enhancement. Lore et al. [36] proposed a superimposed sparse denoising autoencoder (SSDA). This method can adaptively enhance the lightness of natural low-light images, and recognizes the characteristics of the images. Chen et al. [37] established a short exposure and low light image dataset, including the corresponding long exposure reference image. Using this dataset, they developed a dark image enhancement method based on the end-to-end full convolution network. Zhang et al. [38] proposed a deep network for kindling the darkness (KinD), which trained multiple exposure image pairs to estimate the appropriate illumination and reflectance. Lim and Kim [39] proposed a deep stacked Laplacian restorer (DSLR) that uses different levels of Laplacian pyramids to guide the encoder in restoring local details and global illumination. Deep learning has driven advances in image enhancement but the disadvantage is that it requires very high computing power on the device, and these algorithms are usually time consuming. For small devices such as mobile phones, faster computing methods need to be explored.

The exposure fusion method fuses two or more exposure images to obtain a visually harmonious image. Li et al. [40] proposed a multimode image fusion method based on guided filtering, which accelerates the fusion process. The disadvantage is that the two images with little indifference exposure will have severe color distortion after fusion. Fu et al. [41] used a guided filter to estimate the illumination and then designed weights to fuse the lightness-improved and contrast-enhanced images. Ying et al. [42] proposed an exposure fusion framework based on a camera response model and light estimation. A well-exposed fused image is obtained with appropriate exposure and a fusion weight was designed using light estimation techniques. Singh et al. [43] proposed a fusion framework to fuse the channel results of piecewise gamma correction and histogram equalization and used particle swarm optimization (PSO) to find the appropriate multichannel fusion coefficients. The simplicity of the fusion framework is sought, but the inefficiency of the PSO algorithm limits its ability to perform real-time computations. Chang et al. [44] proposed a long-short-exposure fusion network (LSFNet) to solve the problems caused by fusion, such as high noise, motion blur and color distortion.

In this paper, we propose a fast and effective method for low-light image enhancement. The proposed solution has three main contributions.

1) We analyze the widespread phenomenon of cell vibration and establish an energy model of the stimulus intensity by using the mechanism of vibration, which quantitatively describes the relationship between stimulus intensity and energy in the process of cell photothermal transformation.

2) We propose a new global lightness enhancement model that integrates the energy model with the gamma correction model, which improves the deficiency of the gamma correction model. Then, we find the statistical property of lightness and gamma intensity that both the proposed model and the traditional gamma model have, which is helpful to automatically adjust the gamma intensity parameters.

3) We propose a local fusion strategy to improve the results of global lightness enhancements and to restore the details of the defect areas.

The remainder of the paper is organized as follows: In Section II, we fully discuss the proposed energy model, global lightness enhancement methods, and local fusion strategy. Section III shows various experimental analyses and results. Our conclusions are given in Section IV.

II. METHODOLOGY

A. Energy Model

Light transfers energy by independent photons [45], and photoelectric and photothermal effects occur when the photons are absorbed by retinal cells. Energy is converted into electricity and heat. This triggers the movement of ions on the retinal cells, which rely on electrical signals generated by ion movements to produce images [46]. This paper focuses on the cell changes caused by the photothermal conversion process and attempts to quantify the portion of the energy that is converted to heat.

Retinal cells consist of elastic cell membranes and liquid cytoplasm. When photons are absorbed by the retinal cell, energy is transferred from photons to the retinal cell, causing the retinal cell temperature to rise. Therefore, the liquid cytoplasm expands when heated, and the elastic cell membrane expands synchronously. Without subsequent energy supplies, the volume of liquid cytoplasm shrinks, and the cell membrane returns to its initial undisturbed state after the cell temperature drops. Based on this dynamic behavior of expansion and contraction, a free vibration model of a damped system is constructed using mechanical vibration theory. Correspondingly, cell membranes...
with elasticity are considered elastic elements, and liquid cytoplasm with energy absorption and shock reduction are considered damping elements. Because photons cause motion, photons are considered objects, and the stimulus intensity of photons on retinal cells is considered mass. The ability of cells to generate nano-vibrations as a result of heat is ubiquitous and considered a basis for the existence of life signals [47], [48].

Assuming the amplitude of cell vibration is \( s(t) \), based on the theory of mechanical vibration, we assume that cell membranes that can expand and shrink dynamically are elastic elements with a stiffness of \( k \), and the cytoplasm that can absorb energy and reduce vibration is a damper with a viscous damping coefficient of \( c \). The stimulation intensity caused by a single photon is set as the equivalent mass of weight \( M \). Therefore, the displacement \( s(t) \) produced by a single vibration of the cell membrane is:

\[
\ddot{s}(t) + 2\beta \dot{s}(t) + \omega_0^2 s(t) = 0
\]  

where \( \omega_0 = \frac{k}{M} \), \( \beta = \frac{c}{M} \), and \( M > 0 \). \( \beta \) is a viscous damping factor or damping rate, which is a dimensionless parameter. \( \omega_0 \) is the natural frequency of the system. When the stimulus intensity is determined, the natural frequency is only determined by the parameters of the system itself, which is independent of the external stimulus, initial conditions, etc. In addition to the current stimulus intensity. For the selection of a general solution about \( s(t) \), we consider the case of critical damping (\( \beta = \omega_0 \)).

If there is no subsequent energy supplement, the retinal cells will return to the initial stability due to energy dissipation. In this case, the system vibrates only once and returns to the static balance position in a short time. The solution that satisfies the process of cell vibration is:

\[
s(t) = e^{-\beta t} (c_1 + c_2 t)
\]

where \( c_1 \) and \( c_2 \) are two integral constants determined by the starting condition of a motion. Because the energy carried by a single photon currently absorbed is the constant, which means that the stimulus intensity is constant over time, that is \( \frac{dM}{dt} = 0 \), the energy required for the vibration process is:

\[
\varepsilon = \int_0^{+\infty} \left( v^2 \frac{dM}{dt} + M \frac{dv}{dt} \right) dt = \frac{1}{2} M v^2 \bigg|_0^{+\infty}
\]

where, because velocity \( v \) and displacement \( s \) have a differential relationship, i.e.

\[
v = \frac{ds}{dt} = c_2 e^{-\beta t} - \beta e^{-\beta t} (c_1 + c_2 t)
\]

Thus, (3) and (4) can help us derive the expression of \( \varepsilon \).

\[
\varepsilon = c_1 c_2 \sqrt{kM} - \frac{1}{2} c_2^2 M - \frac{1}{2} k c_1^2
\]

(5) indicates that if there is no light stimulation, i.e. \( M = 0 \), there is still a negative energy term related to the membrane stiffness \( k \). Disassembling (5) into a combination of two items, we have the following equation:

\[
\varepsilon = \varepsilon_s - \varepsilon_r
\]

**Definition 1 (Repulsive Energy):** During cell vibration, the energy used by the cell system to reject the stimulus is

\[
\varepsilon_r = \frac{1}{2} k c_1^2
\]

where the minus sign indicates that the cell exhibits spontaneous resistance or repulsion during energy absorption and vibration. This is consistent with the medical observation that live cells are able to respond to external stimuli.

**Definition 2 (Single Stimulation Energy):** When a single photon stimulates a retinal cell, the energy that cells can perceive by its membrane vibration behavior is:

\[
\varepsilon_s = c_1 c_2 \sqrt{k M} - \frac{1}{2} c_2^2 M
\]

Considering the vibration frequency \( f \) of the cell membrane, the energy of the photon flow per unit cycle is:

\[
E = f \varepsilon_s = \frac{\omega_0}{2\pi} e_s
\]

Therefore, it is easy to deduce the expression of \( E \).

**Definition 3 (Cycle Stimulation Energy):** The energy response of retinal cells to photon flow stimulation during the unit cycle is:

\[
E = \frac{1}{2\pi} c_1 c_2 k - \frac{1}{4\pi} c_2^2 \sqrt{k M}
\]

After analyzing the cell membrane vibration, we derive an explicit model consisting of (8) and (10), which measure part of the energy that retinal cells transformed by the vibration effect when stimulated by photons. Next, we will discuss how the proposed energy model works in the field of low-light image enhancement.

### B. Global Lightness Enhancement

Assume that dim areas of the low-light image are marked \( \Omega \), pixel \( P \in \Omega \) and its value is \( I \), \( I \neq 0 \), the lightness mapping function is \( \Psi(\cdot) \), then the mapping function for the constraint of the lightness enhancement process satisfies:

\[
\Psi(I) > I
\]

For (8) and (10), we rewrite them as:

\[
\varepsilon_s = \frac{\lambda^2 \sqrt{\lambda I + (1 - \lambda) I}}{\sqrt{\lambda - 1} - \sqrt{\lambda}}
\]

(12)

\[
E = \frac{\lambda^2}{\sqrt{\lambda - 1} - \sqrt{\lambda}}
\]

(13)

where \( I \in [0, 1] \) and \( \lambda \in [1, 2] \). The condition for writing is \( \varepsilon_s(I = 0) = 0, \varepsilon_s(I = 1) = 1, c_1 = \frac{1}{4\pi^2}, c_2 = \sqrt{2(\lambda - 1)}, k = \frac{4\pi^2}{\lambda - 1}, M = I \). The derivation of (12) and (13) can be found in Appendix A. It is easy to know that (12) satisfies the constraint, allowing it to perform lightness enhancement. The motivation for applying the energy model to the image field is that we assign a retinal cell with the same parameters at each pixel location in the image, and the pixel values are considered to be stimulus intensity signals (\( M = I \)). Therefore, the signal is transformed into a nonlinear energy signal. This is similar to the human visual system. In this paper,
the parameters $c_2$ and $k$ are replaced by the parameter $\lambda$, which can be called the joint factor. It combines the roles of the parameters $c_2$ and $k$, as well as $\sqrt{I}$ and $I$.

In addition, the legacy gamma correction model [49] also satisfies the constraints. In the past, gamma correction models were often used to mimic the nonlinear working mechanism of human vision and display devices. For example, it is often used to correctly reproduce the luminance of CRT displays [50]. However, there is a distinct defect in the gamma correction model, that is, the image is easily fogged. Zhu et al. [51] used this defect feature to generate multiple exposures of a hazy image to obtain a clear image. Since $\varepsilon(\lambda, I)$ is less capable of mapping nonlinearity than the gamma correction model, $E(\lambda, I)$ has a curve property that decreases with increasing $I$. Therefore, a modified gamma correction model combining the energy model is proposed as a lightness enhancement mapping function. The proposed model is as follows:

$$C = \frac{I^\Gamma E(\lambda, I)}{I_{\max}^\Gamma E(\lambda, I_{\max})},$$  

where $C$ is the output of the modified model, and $\Gamma$ is the gamma intensity, which is used to expand the nonlinear mapping space. We have improved the above traditional gamma correction models by fusing the two models. $E$ is inversely proportional to $I$. When $I$ is larger, the value of $E$ is smaller. $E(\lambda, I)$ can be thought of as an adjustment function where the areas with low pixel values are assigned a high product factor while the areas with high pixel values have small coefficients, and then simple normalization measures are used to ensure that the mapping domain falls in the range of $[0,1]$. This approach readjusts the mapping space of the gamma correction function and compresses the increase in the high value range.

Such a function can be constructed based on empirical evidence, for example, $H(x+y+k)$. However, appropriate system parameters are very difficult to determine in such a way. Fig. 2 shows the advantages of the proposed model, which proves that this technique is effective.

Further studies have found a new statistical property associated with the proposed gamma model. Detailed descriptions are as follows:

Statistical Property: For the V-channel in the HSV space, we assume that the average lightness of the nonzero elements in the input image is $I_{vm}$, the average lightness of the nonzero elements in the output image is $C_{vm}$, and the lightness difference is $\Delta V = C_{vm} - I_{vm}$. Given $\lambda$, there is a transformation between the lightness difference and the gamma intensity, and the mathematical model is represented as:

$$\Delta V = c + \frac{1}{a\Gamma + b}$$  

$$C_{vm} = \frac{1}{N - N_0} \sum C_L(x, y)$$  

$$I_{vm} = \frac{1}{N - N_0} \sum I_L(x, y)$$

where the curve parameters ($a$, $b$, and $c$) are different for different input images. $C_L$ is the V-channel intensity of the output image $C$ in HSV space, and $I_L$ is the V-channel intensity of the input image in HSV space. $C_L(x, y)$ and $I_L(x, y)$ are the lightness values at coordinates $(x, y)$. $N$ is the total number of pixels in image $I_L$, and $N_0$ is the number of pixels with a lightness value of 0.

Given the lightness difference $\Delta V$ between the output image and the input image, the corresponding gamma intensity can be recovered according to (15), which adjusts the lightness of the low-light image.

Because different images have different curve parameters, if different images need to achieve the same lightness difference, the gamma intensity needs to be adjusted for each image. Based on statistical properties, this adjustment has been automated to avoid tedious manual tuning. This is very similar to the dynamic lightness regulation by retinal cells.

It should be noted that (15) is an empirical formula based on statistical properties, this adjustment has been automated to avoid tedious manual tuning. This is very similar to the dynamic lightness regulation by retinal cells.

Perception Strategy: For the V-channel in the HSV space, given $\lambda$, we preset a series of gamma intensities, such as sequence $\Gamma = [\Gamma_1, \Gamma_2, \ldots, \Gamma_n]$. For each gamma intensity, the
input image is calculated by (14) to obtain n copies of the output image. Then, the average lightness \( c_{vm} \) of nonzero elements in each output image copy is calculated, and a series of lightness differences \( \Delta V_x = [\Delta V_1, \Delta V_2, \ldots, \Delta V_n] \) can be obtained. By using sequence \( \Gamma_x \), sequence \( \Delta V_x \) and considering efficiency, we tend to use a nonlinear least square method called the trust region reflection algorithm [52], [53] to render the parameters in the curve model.

Based on the above analysis, we summarize the entire framework for global lightness enhancement. Fig. 1 graphically depicts the entire scheme.

**Global Lightness Enhancement Scheme:** For the input low-light image, given \( \lambda \) and gamma intensity sequence \( \Gamma = [\Gamma_1, \Gamma_2, \ldots, \Gamma_n] \), we assume that the lightness difference between the output and the input images is \( \Delta V^\ast \). The scheme consists of two phases. The first phase is to obtain the curve parameters, and the second phase is to enhance the global lightness.

In the first phase, we first obtain the V-channel of the low light image and then use the perception strategy to obtain the lightness difference sequence \( \Delta V_x \) and the three parameter values in the statistical property curve. After obtaining the parameters of the statistical property curve, we can see how the lightness of the global light changes with the gamma intensity.

In the second phase, to obtain a global enhanced image \( C \) with a lightness difference of \( \Delta V^\ast \), we first use the statistical property curve of the first phase to obtain the gamma intensity \( \Gamma \) at \( \Delta V = \Delta V^\ast \). In this way, a special mapping curve of (14) at \( \Gamma = \Gamma^\ast \) is obtained. Finally, this special mapping curve is used to calculate each channel of the low light image in the RGB space to obtain the global enhanced image \( C \).

Fig. 2 shows a series of global lightness enhancement images obtained with the lightness difference \( \Delta V^\ast \) constraint and the statistical property curves of the two models. Traditional gamma correction requires a smaller gamma intensity to achieve the same lightness enhancement. This exacerbates the atomization effect, resulting in a significant decrease in image saturation, and the local features begin to become indistinguishable or even completely lost. After we have applied the proposed energy model, image degradation is reduced, and many features are retained with significant brightness enhancement. The statistical property curve of the proposed model also shows that large lightness enhancement can be achieved without requiring a small gamma intensity.

We proposed a new global lightness enhancement model to de-expose the low-light areas and improve the drawbacks of traditional gamma correction models. At the same time, we noticed that the proposed global model could not handle high-light areas. Next, we propose a local fusion strategy based on fast-guided filtering to improve the local details of the global image.

### C. Local Fusion Strategy

In this subsection, we propose a simple, fast and effective local fusion strategy to further optimize the global image. The proposed strategy consists of three phases: segmentation, filtering and fusion. First, we use segmentation to obtain a mask image (also known as a rough weight map), which depicts the general distribution of light and dark areas in the input image. Then, we use a fast guided filter to obtain fine fusion weights, which can restore the local spatial manifold of the mask image and transform a hard boundary into a soft boundary. Finally, the defective features in global lightness enhancement images are replaced by mask fusion. A detailed description is as follows:

1) **Segmentation:** For the V-channel of the input image, down-sampling is performed first, and then the downsampled image is segmented using a hard threshold to obtain a rough weight map. The segmentation rule we used is:

\[
W_r(x, y) = \begin{cases} 
1, & \text{if } I_v(x, y) < T \\
0, & \text{else} 
\end{cases}
\]

where \( W_r \) is the rough weight map, \( I_v \) is the V-channel downsampled result of the input low-light image, and the down-sampling rate is \( R \), which is used to reduce the amount of data in the segmentation process. \( T \) is the segmentation threshold. \( W_r(x, y) \) and \( I_v(x, y) \) represent the values of \( W_r \) and \( I_v \) at coordinates \( (x, y) \), respectively. Since the effective range of the luminance space is \([0,1]\), we set \( T = 0.5 \) to divide the whole space into two subspaces, light and dark. Areas with values greater than \( T \) are considered light areas, while a coarse weight of 0 is assigned. When the threshold \( T \) is set to a different value (not 0.5), the direct consequence is that the scale range of light and dark areas is no longer equal, which results in a biased information perception. Therefore, threshold \( T = 0.5 \) is a fair solution.

2) **Filtering:** The V-channel downsampled result of the global lightness enhancement image is the guidance image. The coarse weight graph is filtered by a fast guided filter, and the filter results are upsampled to obtain a fine weight map. The filtering process can be simplified as:

\[
W_f = G_f(C_v, W_r, w, r, \eta)
\]

Here, \( G_f(\bullet) \) represents the fast guided filtering process. \( C_v \) acts as a guidance image to correct the local spatial manifold of the rough weight map \( W_r \). \( w \) is the radius of the square filter’s window, \( r \) is the down sampling rate of the fast guided filter and \( \eta \) is a custom regularization constant, usually \( \eta = 0.04 \). The fast guided filter is a classic and rapid computation filter [54], [55]. Finally, the fine weight map is obtained after an upsampling operation with a sampling rate \( R \) is performed on the filtered results \( W_f \).

If there is no region larger than the segmentation threshold in the image, the image is regarded as a dark partition globally, but the spatial local manifold can still be found, and the appropriate fusion weight can be obtained by the fast guided filter. The image \( X_1 \) shown in Fig. 11 is an example of this case. Although there is no light partition exceeding the threshold, our system can still work normally, and the resulting image will not have image quality problems such as unsmoothed artifacts.

3) **Fusion:** The output of the global lightness enhancement scheme is fused with the input image to form a harmonious and
natural image. The local fusion formula is:

\[
F_i = W_f \cdot C_i + (1 - W_f) \cdot X_i
\]

\[
= W_f \cdot \frac{X_i^{\Gamma^*} \cdot E(\lambda, X_i)}{X_{i_{\text{max}}}^{\Gamma^*} \cdot E(\lambda, X_{i_{\text{max}}})} + (1 - W_f) \cdot X_i
\]

(20)

where \(X\) is the input image, \(F\) is the output image, and \(i\) represents the \(i\)th channel in the RGB space. \(\cdot\) is the elementwise product, and \(X_i^{\Gamma^*}\) represents the result of the gamma transformation of \(X_i\) when the gamma intensity is \(\Gamma^*\). Fig. 3 is a typical example of how a local fusion strategy works with the proposed modified model to achieve lightness enhancement. The sky of the input image \(X\) in Fig. 3 seems clean. After enhancing lightness, details appear in the dark, but the sky details are lost. By using the local fusion strategy, the sky features in the input image are successfully fused, resulting in clear and well-balanced images.

De-masking is a direct and effective remedy solution, which has a significant advantage over the iterative solutions on with low computational cost. Subsequent experiments have confirmed this analysis.

**D. Discussion**

*Why replace the parameters in the energy model with parameter \(\lambda\)?* The parameters \(c_1, c_2,\) and \(k\) are used to describe the movement and intrinsic properties of real retinal cells. The values of \(c_1, c_2,\) and \(k\) need to be determined in physical measurement experiments. For different types of cells, we assume that the values of these parameters are different. Therefore, to use the proposed energy model, we used the parameter \(\lambda\) to rewrite the equation.

*How large should the filter window be?* The window radius \(w\) is related to the amount of local detail in the weight map \(W_f\). As the window size increases, the local feature information perceived by the window increases, which provides guidance for correcting local spatial manifolds in the rough weight maps \(W_f\). To obtain a fine weighted map with rich local details, through experimental evaluation, the window radius \(w\) we select is

\[
w = \left\lfloor \frac{\min(\text{row}, \text{col}) - 1}{2} \right\rfloor
\]

(21)

\(\lfloor \cdot \rfloor\) is a downward rounding operation to avoid the window size exceeding the limit. \(\text{row}\) represents the number of rows of image \(C_v\), and \(\text{col}\) represents the number of columns of image \(C_v\).

*How does a coarse or fine weight map affect the final fusion result?* Fig. 4 shows the fusion results using coarse and fine weight maps. Rough weight maps produce disharmonious fusion results because local manifold characteristics are not taken into account. With the help of the fast guided filtering method, an estimation solution of the local spatial manifold structure is found, which successfully solves the problem caused by rough fusion and makes the fused image appear natural and harmonious.

**III. Experimental Work**

In this section, we first examine the impacts of the involved parameters. Next, we present statistical evidence supporting (15). Afterwards, we compare our method with several state-of-the-art methods from the perspective of subjective qualitative and objective quantitative evaluation. Finally, we give the time consumption of different enhancement methods. The comparison methods include HE [2], NPEA [17], SRIE [18], LIME [9], AIEMC [29], ASTF [26], MF [41], EFF [42], and KinD [38]. A test platform based on a computer (AMD Ryzen 5 2400 G @3.6 GHz, 16 GB RAM and Windows 10 operating system) is established. The software was developed with MATLAB 2018b. To compare the methods, we use the MATLAB code published by the authors. The KinD PyTorch code is provided by the author, and the NVIDIA GeForce RTX 3060 (6 GB) graphics card is used to speed up the calculation. The images used in the experiment come from publicly accessible datasets. The majority
of our experimental images are from four databases, SIDD [56], [58], the ESPL-LIVE HDR Image Quality Database (HDR) [59], [60], ExDark [61], NASA, LOL [62] and a synthetic dataset [62].

A. Parameters’ Impact

The parameters to be configured in the global lightness enhancement stream are $\lambda$, sequence $\Gamma_s$, and $\Delta V^*$. The parameters to be discussed in the local fusion strategy are $R$, $r$ and $w$. Through experimental analysis, a set of appropriate parameters are configured as $\lambda = 2$, $\Gamma_s = [0.3, 0.8, 1.3, 1.8]$, $\Delta V^* = 0.25$, $R = 2$, and $r = 10$. The window radius $w$ is shown in (21). This set of parameters is used for subsequent qualitative evaluation experiments and objective quantitative examinations. Next, we will discuss their impact separately.

The parameter $\lambda$ is configured as 1.1, 1.5 and 2. The other parameters are unchanged. The range of parameter $\lambda$ is (1,2). Because there are infinite parameter choices in a finite space, to demonstrate the system performance under different parameter values, we only show three possible cases, i.e., $\lambda = 1.1, 1.5$ and 2. The results are shown in Fig. 5. The visual performance of the fusion results is satisfactory, and the difference is very small, except that the fusion results are gray when $\lambda = 1.1$. A closer examination shows that the colors are more vivid when $\lambda = 2$. More detailed numerical results are shown in Fig. 5(e).

We use three statistical states, namely, the mean difference of lightness $\Delta V_m$, mean difference of saturation $\Delta S_m$ and difference of lightness and saturation $D_m$, to analyze the effect of parameter $\lambda$. The three statistical states are calculated as follows.

$$\begin{align*}
\Delta V_m &= F_{vm} - I_{vm} \\
\Delta S_m &= I_{sm} - F_{sm} \\
D_m &= F_{vm} - F_{sm}
\end{align*}$$

where $F_{vm}$ and $F_{sm}$ are the average lightness and saturation of the fused image, respectively. $I_{sm}$ is the average saturation of a low-light image. Fig. 5(e) shows that $\Delta V_m$ has decreased from 0.25 to approximately 0.2 due to the low illumination of the low-light image. Because the difference between the three $\lambda$ is small, this means that $\lambda$ has a slight effect on $\Delta V_m$. The color attenuation prior proposed by Zhu [63] indicates that there is a saturation decrease, and $D_m$ increases during the transition from a clear image to a hazy image. Therefore, $\Delta S_m$ and $D_m$ of a clear and vivid fusion image are at a low level. Since $\Delta S_m$ and $D_m$ decrease significantly when $\lambda = 2$, we choose $\lambda = 2$ based on the above analysis.

The gamma sequence $\Gamma_s$ is a reference element of curve fitting. It represents a combination of gamma distributions used to obtain the three parameters of (15). To solve the parameters of the nonlinear model shown in (15), at least four reference coordinate points are required. We suggest that the range of preselected points for $\Gamma_s$ be [0.3, 2.2]. For (14), if $\lambda = 2$ and the maximum value is obtained at $I = 1$, the variation of $C$ with the gamma intensity can be obtained as shown in Fig. 6.

Since the combinatorial number of gamma sequences is $C_4^2 (n \rightarrow \infty)$, we only discuss and compare the following four gamma sequences for analysis purposes.

$S_1 : \Gamma_s = [0.3, 0.8, 1.3, 1.8] ; S_2 : \Gamma_s = [0.8, 0.9, 1.1, 1.2] ;
S_3 : \Gamma_s = [0.3, 0.5, 0.6, 0.8] ; S_4 : \Gamma_s = [1.3, 1.5, 1.6, 1.8] .

We continue to use Fig. 5(a) in the experiments. Fig. 7 shows the fitting and the mean square error (MSE) of the four sequences. In these four cases, the sequence $S_3$ has the lowest fitting error, and the sequence $S_4$ has the worst fit. Therefore, we recommend the gamma sequence $S_1$.

The lightness difference $\Delta V^*$ is an important parameter. Because $C_{vm} = I_{vm} + \Delta V^*$, the lightness of the global lightness enhancement image changes with $\Delta V^*$ and ultimately affects the lightness of the fused image. For low-light image enhancement, $\Delta V^*$ must be greater than 0. To select an appropriate lightness difference, we calculated the average lightness distribution of the four databases, and the results are shown in Fig. 8. Details are ExDark: 7363 images; HDR: 250 pairs of images (including low-light and normal); NASA: 22 images; SIDD: 74 images. The Exdark dataset contains 12 categories of images. These categories are bicycle, boat, bottle, bus, car, cat, chair, cup, dog, motorcycle, people and table. The numbers of these
categories are 652, 679, 547, 527, 638, 735, 648, 519, 801, 503, 609 and 505.

The statistical results show that the average lightness of most of the low-light images is less than 0.4, the median is approximately 0.2, and the average upper and lower bounds are 0.0623 ~ 0.4724. Most of the normal illuminated images fall between 0.4441 and 0.5776, and the upper and lower bounds are 0.29454 ~ 0.77257. It is recommended that $\Delta V^*$ be set between $[0.2, 0.3]$. This conclusion is consistent with the observation in Fig. 2; that is, the lightness enhancement of the image is insufficient when $\Delta V^*$ is less than 0.2, and it may be too bright when $\Delta V^*$ is greater than 0.3. In addition, it is assumed that for different images, $\Delta V^*$ conforms to a mathematical distribution within the range [0.2, 0.3], such as a normal distribution. Therefore, based on the above analysis, we recommend $\Delta V^* = 0.25$.

Downsampling rates $R$ and $r$ are set up to speed up the algorithm. To speed up the acquisition of rough weight maps, we downsample the V-channel of the input image. On this basis, a fast guided filter is used to further accelerate the calculation. The downsampling rate $r$ of the fast-guide filter does not significantly change the filtering result. Therefore, we set $r = 10$. Fig. 9 shows the result of the weight map generated by different combinations of downsampling rates $R$ and $r$. Obviously, $R$ significantly affects the local details of the fine weight map. When $R = 2$, there is a small loss of detail in the weight map, and an increase in the NIQE score also means a decrease in image quality. However, the difference between the fusion results of the four combinations is very small, and it is difficult for human eyes to detect the difference. Therefore, we recommend setting $R = 2$ to speed up the algorithm calculation while allowing the weight map to have a small loss of detail.

For the window radius $w$, we do not recommend a fixed window size but use a dynamic window size for different images. This is because the spatial distribution range of image features changes with the size of the image, which makes the small filter window in the small-scale image unable to adapt to the large features in the large-scale image. Therefore, due to the incomplete collection of local spatial manifold information, some spatial manifold information is lost, and finally, the local manifold of the weight map is not sound. Considering that the larger the window is, the more characteristic information it will collect, we use the dynamic window radius shown in (21). Fig. 10 shows the weight map generated by two different sizes of images at different window sizes. It is easy to know that the dynamic window

![Fig. 9. Weight maps and NIQE scores for different down sampling rates. (a) Top: Input image with uneven illumination from the Exdark dataset. Bottom: NIQE scores for different down sampling rates. (b)~(e) The result at $(R, r) = (1,1), (2,1), (1,10) and (2,10)$, respectively. From top to bottom are the fusion results and weight maps. The black border is artificially added to highlight the boundary. The NIQE scores of the four combinations differ slightly, with a maximum deviation of approximately 0.11. The NIQE score shows that the influence of the downsampling rate $R$ on the images is more significant than that of the downsampling rate $r$.](image)

![Fig. 10. Weight maps and entropy for different window radii. The dimensions from top to bottom are $375 \times 500 \times 3$ and $2000 \times 1312 \times 3$, respectively. (a)~(c) Weight maps are generated with 32, 64, and 128 filter radii, respectively. (d) The weight map is generated by using a dynamic window radius. (e) Entropy of four window radii. Top: Image “group photo”. Bottom: Image “smiling woman”. The entropy metric [64] quantitatively describes the impact of different window sizes on the weight map. The black border is artificially added to highlight the boundary.](image)
radii are 187 and 655, respectively. When the window radius is 32, only a few local spatial manifold structures are found. With the increase in the window radius, the local space manifold becomes stable gradually and reaches the final state at the dynamic window radius. In addition, increasing entropy also means that more manifold structural information is found. The fused result has a flaw similar to Fig. 4 when the window is small, so there is no additional display of the result image to save space. Similarly, the fusion results of the dynamic window are shown in Fig. 5(d) and Fig. 9(e).

### B. Statistical Cases

The Exdark dataset is used to verify the validity of the statistical property and to illustrate the rationality of (15). Specifically, we examined a total of 7363 images from 12 categories of the Exdark dataset. For any image, the sequence $S_1$ is used to obtain the curve parameters. Then, we calculated the MSE of the gamma interval [0.3, 2.2] with a step size of 0.05. The maximum, minimum and mean values in the MSE for each category and for the entire dataset are recorded in Table I. It should be noted that the MSE metric is only used to verify the rationality of (15). For the proposed model (14) and the gamma correction model, the fitting error on the whole dataset is very small, which indicates that for a natural low-light image, using the proposed model or gamma correction, a series of results can form a curve as described by the statistical property. It also shows that the mathematical expression of (15) is reasonable. The proposed method has a slightly larger MSE value than the gamma correction model, with MSE values ranging from $10^{-4}$ to $10^{-8}$. This result proves the validity of (15) with a low MSE indicator score. However, the statistical property is not always valid for a local block of the image. If the element value of a local block is 1, it remains the same throughout the process. Therefore, if this local block is treated as an image separately, it does not have statistical properties.

### C. Subjective and Qualitative Evaluations

We selected images from the SIDD, Exdark, HDR, LOL and synthetic datasets as references. Figs. 11, 12 and 13 show the results of the nine algorithms. Image $X_1$ is a color correction board in a low-light state, which can be used to compare the color recovery capabilities of different algorithms. Our method yields clear and appropriate colors, and other methods result in a distorted appearance, i.e. white squares are significantly reddish or gray. This indicates that the results of other methods are biased due to the compression or expansion of the grayscale. For image $X_1$, the image size is $5312 \times 2988 \times 3$. The KinD method cannot deliver the calculation due to the limitation of the video memory, so we downsample the image to $960 \times 540 \times 3$. However, the
result of the KinD method does not obtain a significant lightness improvement, and uneven color patches appear. For image X2, our results are very similar to EFF. Local facial details and colors show that the results of our method and EFF are more in line with human observations. The KinD method significantly increases the lightness level, but local texture features are poorly handled, such as blurred license plate numbers and large area artifacts in the sky. Other methods have obvious redness or overexposure.

Fig. 12 shows the enhancement results of a low-light image with noise. It can be seen from the cat’s ear that HE and AIEMC overemphasize the local details and lose the features. The HE method hardly amplifies the noise in the shadow area on the right side of the cat’s ear, while the other methods amplify the noise signal, the worst are NPEA, AIEMC, MF and KinD. For the KinD method, there are serious artifact defects near the grass, and the noise in the cat’s ear shadow area is overamplified. The LIME method produces appropriate results. The overall color saturation is high, but there is still strong noise. The noise intensity of our method and EFF is low, and the noise signal is not overamplified.

The six sets of legends in Fig. 13 are taken from the HDR, LOL and synthetic datasets, where images in normal light are regarded as “ground truth” and can be used to reveal the appropriateness of color and texture detail. For a series of results from image X4, the proposed method and EFF are very close to reality. For image X5, HE produces significant color distortion and loses some of the features of the ceiling lamp. KinD also loses the local texture of the ceiling lamp to make it no longer visible. NPEA and LIME also produced a discordant result, which excessively enhanced the ceiling lamp area. SRIE and KinD produce additional shadows in the suspension area. The remaining methods deal with the characteristics of the ceiling lamp area better. In addition, it can be seen that our method fully restores the details, and the lightness is appropriate. For image X6, the result obtained by our method is very close to the reference image. Most of the contrast methods retain the color saturation of the sky in the low light image but some local textures are lost. HE and KinD achieve poor results.

For image X7, our results present milky white square boxes, which are different from the others. For image X8, our results are close to those of normal images but the NPEA method appears to be superior. For image X9, the lightness difference between the normal image and the test image is very small, but our method still achieves suitable results.

For the nine comparison methods, the lightness level is not an adjustable parameter, and there is no lightness relationship similar to (15) in this paper. Eq. (15) enables our system to control lightness levels. After setting a specified lightness level or lightness difference, the gamma value parameters can be automatically adjusted for different images based on (15). The comparison method lacks the function of flexibly adjusting and
Fig. 13. Comparison results for image examples. Images X_4, X_5, X_6 and normal images are taken from the HDR dataset. Image X_7 is taken from the LOL dataset. Images X_8 and X_9 are taken from the synthetic dataset. Normal represents the image under normal illumination.
controlling the lightness level so that the lightness level of the result becomes uncontrollable.

A series of image comparison results show that the proposed method has better adaptability to various complex low-light environments and can reveal the dark area details of the low-light image and reproduce natural colors.

D. Objective Quantitative Examination

In this section, we show the index scores of different images in the qualitative evaluation experiment. In addition, to objectively evaluate the performance of different algorithms, the HDR, LOL and synthetic datasets are used for objective and quantitative examination. These three datasets have reference images under normal lighting, and performance measurements using these datasets are more convincing than using datasets without reference objects (such as SIDD, ExDark and NASA).

A total of 250 pairs of images are selected from the HDR dataset to undertake the index evaluation experiment. The HDR dataset uses 605 scenes, totaling 1811 images. After carefully examining each image, we found that there are many pairs of images in the HDR dataset with inconsistent sizes, artifacts and other defects, as shown in Fig. 14. To improve the reliability of the experimental data and results, we manually browse through each image and eventually filter out 250 pairs of visually pleasing images. These 250 pairs of images have been packaged into a new modified dataset and published on the Internet. The LOL dataset contains 500 pairs of low/high light images, and the synthetic dataset contains 1000 pairs of low/high light images. Considering that the KinD algorithm has used the LOL dataset as the training set to obtain the network parameters, to fairly evaluate the performance of the algorithm, the index score of the LOL dataset does not include that of the KinD algorithm. The index evaluation experiment consists of 7 different measurement methods, which are derived from two different index systems of no-reference assessment and referential assessment. These are NIQE [65], BRISQUE [66], ΔE, PSNR, MSSIM [67], VIF [68] and LOE [17]. Among them, the smaller the index score is, the higher the image quality represented by NIQE, BRISQUE, ΔE and LOE. The other three indicators are the opposite.

NIQE and BRISQUE are two popular blind image quality assessment models, both of which are based on a statistical analysis of spatial domain characteristics to avail the evaluation score and depict the distortion intensity of the image.

ΔE is a measure of color accuracy that represents the distance between two colors. This index needs to be calculated in the LAB color space. Specifically, we used the ΔE index to evaluate the color difference between the enhanced image and the reference image. If the index score is smaller, the color representing the enhanced image is closer to the reference image. The ΔE index is calculated from the following formula.

\[
\Delta E = \frac{1}{N_I} \sum_{i,j \in \Omega} \sqrt{(L'_{ij} - L_{ij})^2 + (a'_{ij} - a_{ij})^2 + (b'_{ij} - b_{ij})^2}
\]

(23)

where \(N_I\) is the total number of pixels in an image of size \(m \times n\), \((i, j)\) is the image coordinate number, \((L'_{ij}, a'_{ij}, b'_{ij})\) and \((L_{ij}, a_{ij}, b_{ij})\) are the specific values of the enhanced image and the reference image in the LAB space at coordinates \((i, j)\), respectively.

PSNR, MSSIM, and VIF are three commonly used reference image quality assessment indexes. PSNR is a measure of the similarity between two images. MSSIM is usually used to describe the detailed reconstruction intensity of an image. VIF is the abbreviation of visual information fidelity, which depicts the degree of information fidelity. The LOE index is a measure of the lightness order between an enhanced image and a low-light image, which is used to evaluate the degree of lightness order preservation.

Figs. 15 and 16 show the NIQE and BRISQUE scores of no-reference images \(X_1\), \(X_2\), and \(X_3\), respectively. For image \(X_1\), our method achieves a very pleasant visual effect but does not obtain the best score, which means that these two blind evaluation criteria do not always work well. For image \(X_2\), our NIQE score is almost the same as any other method except ASTF, although our method yields suitable facial results. In addition, our
method achieves almost the best BRISQUE scores on images $X_2$ and $X_3$.

Tables II, III and IV show the scores of different methods under seven indicators and show the quality assessment of low-light images (briefly expressed as Low) and normal light images (briefly expressed as Normal), but these two types of results are only for display. Bold fonts are used to highlight the top algorithms in the tables.

For the HDR dataset, the proposed method achieves the best performance for the five indicators and is only slightly inferior for the VIF and LOE indicators.

For the LOL dataset, our method has the highest scores on the $\Delta E$ and MSSIM metrics and ranks third in the NIQE, BRISQUE and PSNR metrics. For this dataset, the VIF and LOE metrics are slightly lower.

For the synthetic dataset, our method obtains the highest scores on the MSSIM and LOE metrics, ranks second in the BRISQUE and PSNR metrics, and ranks third in the $\Delta E$ metric.

For this dataset, the VIF and NIQE metrics are slightly lower.

Combining the results of the three datasets, we conclude that our method is very competitive.

### E. Time Cost

Another important measure of algorithm performance is the time cost. Table V records the time cost of different enhancement algorithms at different picture sizes. The time cost was obtained by averaging the time of 10 records. We used a GPU to speed up the calculation for the Kind method, while the other methods were executed on a CPU.

![Table V](image.png)

**Note:** The above text contains mathematical notations and tables, which are essential for understanding the performance evaluation of the proposed algorithm. The tables and figures are crucial for interpreting the results, such as the time cost under different image sizes and the average score of different enhancement algorithms under seven metrics for HDR datasets.
use a CPU. Constrained by the video memory of the GPU, we cannot complete the time statistic test of the KinD method when the image size is 2000×1312×3 or above. As shown in Table V, the time cost of the proposed algorithm is the lowest, and the time consumption is reduced by at least half. The EFF method is close to the proposed method on the seven indexes. However, our method is at least 4× more efficient than the EFF method for the images with a size of 2000×1312×3 and more, which means that real-time processing of video with a conventional image size is entirely possible. We implemented the algorithm using MATLAB 2018b, so it can be further accelerated by using faster programming languages, such as C/C++.

IV. CONCLUSION

In this paper, we have proposed a simple yet effective low-light image enhancement method to perceive unknown information from dark areas. This is the first application of the proposed cell vibration model in the field of image vision. Based on the analysis of the energy model, we have improved the standard gamma model, that is, we retained more image features with a large increase in lightness. After further investigation, the statistical properties revealed a relationship between natural image lightness and gamma intensity. To further optimize the local details of global lightness enhancement images, we proposed a local fusion strategy that successfully recovers the nature of the original low-light images. Comprehensive experiments have been undertaken and the results show that our algorithm recovered image details well and effectively avoided overenhancement and color distortion. Compared with several state-of-the-art methods, our method achieves a step forward in image enhancement tasks. In future work, we will continue to optimize the performance and time consumption of the algorithm, hoping to maintain systematic efficiency even for large-scale images of 100 million pixels.

APPENDIX A

EQUATION DERIVATION

This appendix mainly describes the mathematical derivation process of (12) and (13) in subsection B. The stimulation energy equations we derived are as follows. To apply these two equations to the image field, we need to rewrite the equations.

\[ E = \frac{1}{2\pi} c_1 c_2 k - \frac{1}{4\pi} c_2^2 \sqrt{kM} \]  
\[ \varepsilon_s = c_1 c_2 \sqrt{kM} - \frac{1}{2} c_2^2 M \]  

Assume

\[ \begin{cases} \lambda = c_1 c_2 \sqrt{k} \\ 1 - \lambda = -\frac{1}{2} c_2^2 \end{cases} \]

Due to the smoothness of (24), the above assumptions ensure that \( \varepsilon_s(M = 0) = 0, \varepsilon_s(M = 1) = 1. \) We can derive

\[ \begin{cases} k = \frac{\lambda^2}{c_2^2} \\ c_2 = \sqrt{2(\lambda - 1)} \end{cases} \]

Substitute the above parametric relationships into (25),

\[ E = \frac{1}{2\pi} c_1 c_2 k - \frac{1}{4\pi} c_2^2 \sqrt{kM} = \frac{\lambda^2}{2\pi c_1 c_2} - \frac{1}{4\pi} c_2^2 \frac{\lambda}{c_1 c_2} \sqrt{M} = \frac{\lambda^2}{2\pi c_1 c_2} - \frac{1}{4\pi c_1} \sqrt{M} = \frac{\lambda^2}{2\pi c_1 \sqrt{2(\lambda - 1)}} - \frac{\lambda}{4\pi c_1} \sqrt{\frac{M}{\lambda - 1}} = \frac{1}{2\pi \sqrt{\lambda - 1}} \left[ \frac{\lambda^2}{\sqrt{\lambda - 1}} - \lambda \sqrt{\frac{(\lambda - 1)M}{\lambda - 1}} \right] \] \( (28) \)

To make (28) more concisely, we assume

\[ c_1 = \frac{1}{2\sqrt{2\pi}} \] \( (29) \)

Therefore, we can obtain

\[ E = \frac{\lambda^2}{\sqrt{\lambda - 1}} - \lambda \sqrt{(\lambda - 1)M} \] \( (30) \)

It is assumed that the pixel value of an image represents the stimulus intensity, that is, \( M = I \), so

\[ \varepsilon_s = \lambda \sqrt{I} + (1 - \lambda) I \]

\[ E = \frac{\lambda^2}{\sqrt{\lambda - 1}} - \lambda \sqrt{(\lambda - 1)I} \] \( (31) \) \( (32) \)

REFERENCES

[1] R. Chandrasekharan and M. Sasikumar, “Fuzzy transform for contrast enhancement of non-uniform illumination images,” IEEE Signal Process. Lett., vol. 25, no. 6, pp. 813–817, Jun. 2018.
[2] R. Hummel, “Image enhancement by histogram transformation,” Comput. Graph. Image Process., vol. 6, no. 2, pp. 184–195, 1977.
[3] Y.-T. Kim, “Contrast enhancement using brightness preserving histogram equalization,” IEEE Trans. Consum. Electron., vol. 43, no. 1, pp. 1–8, Feb. 1997.
[4] T. Celik and T. Tjahjadi, “Contextual and variational contrast enhancement,” IEEE Trans. Image Process., vol. 20, no. 12, pp. 3431–3441, 2011.
[5] V. Caselles, J. L. Lisani, J.-M. Morel, and G. Sapiro, “Shape preserving local histogram modification,” IEEE Trans. Image Process., vol. 8, no. 2, pp. 220–230, Feb. 1999.
[6] J.-L. Lisani, “An analysis and implementation of the shape preserving local histogram modification algorithm,” Image Process. On Line, vol. 8, pp. 408–434, 2018.
[7] J. Cai, S. Gu, and L. Zhang, “Learning a deep single image contrast enhancer from multi-exposure images,” IEEE Trans. Image Process., vol. 27, no. 4, pp. 2049–2062, Apr. 2018.
[8] W. Kim, R. Lee, M. Park, and S. Lee, “Low-light image enhancement based on maximal diffusion values,” IEEE Access, vol. 7, pp. 129150–129163, 2019.
[9] X. Gao, Y. Li, and H. Ling, “LIME: Low-light image enhancement via illumination map estimation,” IEEE Trans. Image Process., vol. 26, no. 2, pp. 982–993, Feb. 2017.
[10] E. H. Land, “The retinex,” Amer. Sci., vol. 52, pp. 247–264, 1964.
[11] E. H. Land, “An alternative technique for the computation of the designator in the retinex theory of color vision,” Proc. Nat. Acad. Sci., vol. 83, no. 10, pp. 3078–3080, 1986.
[12] J. A. Frankle and J. J. McCann, “Method and apparatus for lightness imaging,” U.S. Patent US04384336A, 1983.
[13] B. Funt, F. Ciurea, and J. McCann, “Retinex in Matlab,” in Proc. Color Imag. Conf., 2000, pp. 112–121.
Xiaozhou Lei received the B.S. degree in mechanical design manufacture and automation major and the M.S. degree in mechatronic engineering from the Wuhan Institute of Technology, Wuhan, China, in 2015 and 2018, respectively. He is currently working toward the Ph.D. degree in control science and engineering with Shanghai University, Shanghai, China. His research interests include machine vision, artificial intelligence, image enhancement, and object detection.

Zixiang Fei received the bachelor’s degree from Liverpool John Moores University, Liverpool, U.K., and the master’s degree from the University of York, York, U.K., and the Ph.D. degree from the University of Strathclyde, Glasgow, U.K., in 2020. He was a Postdoc with the University of Strathclyde, in 2021. He is currently a Lecturer with Shanghai University, Shanghai, China. His major research interests include computer vision, machine learning, object recognition and deep learning.

Wenju Zhou received the B.Sc. and M.Sc. degrees from Shandong Normal University, Jinan, China, in 1990, and 2005, and the Ph.D. degree from Shanghai University, Shanghai, China in 2014. He is currently a Professor and a Doctoral Supervisor with Shanghai University. His research interests include robotics control, machine vision, and the industry applications of automation equipment.

Huiyu Zhou received the B.E. degree in radio technology from the Huazhong University of Science and Technology, Wuhan, China, the M.Sc. degree in biomedical engineering from the University of Dundee, Dundee, U.K., and the D.Phil. degree in computer vision from Heriot-Watt University, Edinburgh, U.K. He is currently a Professor with the School of Computing and Mathematical Sciences, University of Leicester, U.K. He has authored or coauthored widely in the field.

Minrui Fei received the B.S. and M.S. degrees in industrial automation from the Shanghai University of Technology, Shanghai, China, in 1984 and 1992, respectively, and the Ph.D. degree in control theory and control engineering from Shanghai University, Shanghai, China, in 1997. Since 1998, he has been a Full Professor with Shanghai University. He is currently the Chairman of Embedded Instrument and System Subsociety, and the Standing Director of China Instrument & Control Society, Chairman of Life System Modeling and Simulation Subsociety, Vice-Chairman of Intelligent Control and Intelligent Management Subsociety, and the Director of Chinese Artificial Intelligence Association, a Fellow of China Simulation Federation. His research interests include networked control systems, machine vision, artificial intelligence, intelligent control, complex system modeling, hybrid network systems, and field control systems.