



Effective multi-scale enhancement fusion method for low-light images based on interest-area perception OCTM and “pixel healthiness” evaluation

Yi-lun Wang¹ · Yi-zheng Lang¹ · Yun-sheng Qian¹

Accepted: 9 June 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Low-light images suffer from low contrast and low dynamic range. However, most existing single-frame low-light image enhancement algorithms are not good enough in terms of detail preservation and color expression and often have high algorithmic complexity. In this paper, we propose a single-frame low-light image fusion enhancement algorithm based on multi-scale contrast–tone mapping and “pixel healthiness” evaluation. It can adaptively adjust the exposure level of each region according to the principal component in the image and enhance contrast while preserving color and detail expression with low computational complexity. In particular, to find the most appropriate size of the artificial image sequence and the target enhancement range for each image, we propose a multi-scale parameter determination method based on the principal component analysis of the V-channel histogram to obtain the best enhancement while reducing unnecessary computations. In addition, a new “pixel healthiness” evaluation method based on global illuminance and local contrast is proposed for fast and efficient computation of weights for image fusion. Subjective evaluation and objective metrics show that our algorithm performs better than existing single-frame image algorithms and other fusion-based algorithms in enhancement, contrast, color expression, and detail preservation.

Keywords Low-light images · Image enhancement · Multi-scale fusion · Contrast–tone mapping

1 Introduction

Low-slight images are a widespread class of images in the real world, a degradation of image brightness and contrast due to many factors such as insufficient light intensity or shadows [1].

Despite the amazing advances in image sensor technology, longer exposure time when shooting in low-light scenes can also lead to phenomena such as artificial ghosting and partial over-exposure. To enhancing the brightness and contrast of low-light images, particularly to achieve high dynamic range, is a critical part of image pre-processing to meet the growing demand for complex, all-weather scene imaging.

In recent years, numerous image enhancement methods have been developed, encompassing techniques based on histogram equalization, adaptive curve mapping, retinex-based methods, and learning-based methods. Each method has its unique advantages. For instance, histogram equalization-based methods enhance images by redistributing their grayscale values, leveraging the statistical distribution of image brightness for significant enhancement effects with high efficiency [2]. Adaptive curve mapping techniques, on the other hand, enhance images while maintaining a natural appearance, thus achieving excellent visual outcomes [3]. Retinex-based methods, a significant category in image enhancement, have seen rapid advancements, enhancing images by estimating illumination components with remarkable efficacy [4].

These methods perform adequately in uniformly low-light conditions and when the scene has few observation targets. However, in scenarios involving diverse light sources, shadows, or clear foreground, midground, and background elements, they often fail to ensure uniform enhancement across all areas, leading to over- or under-enhanced regions.

✉ Yun-sheng Qian
yshqian@mail.njust.edu.cn

¹ School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China

Fusion-based methods serve as an excellent supplement in these cases [5–7]. They generate multiple artificial images, each enhancing specific areas or grayscale ranges, which are then combined to produce the final image. This method tends to ensure that all regions of the image have been given satisfactory brightness and contrast without losing sight of one or the other.

Nevertheless, traditional fusion-based enhancement strategies have their limitations. First, the fixed number of artificial images can lead to computational redundancy in simple scenes or unsatisfactory enhancement in complex lighting conditions. Second, artificial image generation often overlooks the original light–dark relationships within the image, potentially resulting in unnatural outcomes post-fusion. Third, the fusion step, crucial for enhancement, typically does not prioritize edge and texture preservation, risking detail loss in the output image.

To play the advantages of fusion methods in enhancing low-light images under complex lighting, while addressing these issues, this paper introduces an adaptive multi-scale image enhancement and fusion method for single-frame low-light images. This method adaptively calculates the optimal number of artificial images based on the grayscale distribution characteristics of the image, significantly optimizing the naturalness and detail representation of the fused image. Initially, we employ histogram statistics of the V channels in HSV color space along with principal component peak analysis to adaptively ascertain the scale and target enhancement range of the multi-scale artificial image sequence. Secondly, we improve the traditional OCTM method by making it applicable to multi-scale artificial images. We call this refined version as "interest-area perception OCTM" (IAP-OCTM), which adaptively calculates the parameter based on the grayscale range of the interest-areas in the current artificial image. This process generates multi-scale enhanced image sequences which are ready for subsequent fusion. Thirdly, we propose a new "pixel healthiness" evaluation method based on global luminance and local contrast to calculate the weight of each pixel in each artificial image, which effectively enhances the contrast and luminance and preserves detail. The enhanced image from the proposed method combines brightness enhancement, contrast enhancement, detail protection and optimization of color expression to achieve the best low-light image enhancement.

The main contributions of our method are as follows:

1. This paper proposes a new multi-scale parameter determination method for determining the number of images in multi-scale enhanced artificial images sequence and the grayscale range for interest-areas in each generated image. The method is based on histogram statistics and principal component peak analysis. It can solve the problem of under-enhancement in complex illumination

scenes or computational redundancy in uniform low-light scenes caused by the fixed number of artificial images in traditional fusion enhancement methods.

2. This paper proposes an improved OCTM method called "interest-area perception OCTM" (IAP-OCTM) for enhancing low-light images. The key of this method lies in its ability to adaptively enhance the areas of interest within the current artificial image while also considering the light–dark relationship of these regions in the original image. IAP-OCTM ensures that all major regions of the original low-light image been enhanced, while avoiding unnatural phenomena such as chaotic light–dark relationships in the final enhanced image.
3. A "pixel healthiness" evaluation method based on global illuminance and local contrast is proposed. It integrates the grayscale and edge detail of the pixel. We use this method as a criterion to calculate each pixel's weight. The method is simple and efficient, with good edge protection and excellent detail performance.

2 Related works

With the ongoing advancements in image processing technology, many image enhancement methods based on single-frame low-light images have been proposed. They can be primarily categorized into four main types: value-based, retinex-based, fusion-based, and deep learning-based.

Value-based methods are methods for redistributing the global grayscale distribution of low-light images based on pixel-value statistics, thereby improving the brightness and contrast of the image, including the classic histogram equalization (HE) and gamma correction (GC) methods. Researchers have also proposed various improvements in the early years, such as contrast limited adaptive histogram equalization (CLAHE) [8], adaptive gamma correction with weighting distribution (AGCWD) [9] and optimal contrast-tone mapping (OCTM) [10]. Although these methods can achieve luminance and contrast enhancement, they cannot cope well with noise and uneven illumination effects. The enhancement can be over-enhanced or unnatural. Later, Gu et al. [11] employed visually noticeable regions and Su et al. [12] combined optimal contrast-tone mapping with human eye observation habits to further prevent the generation of unanticipated artifacts and over-enhancement. However, the contrast and edge sharpness of the enhanced image are still unsatisfactory.

Retinex-based methods are one of the most classic approaches in the image enhancement area. Retinex theory was first proposed by Land and McCann [13] in 1977. They proposed that color is determined by the ability of an object to reflect long-, medium- and short-wave light and the

color is not affected by the non-uniformity of light. However, after extensive subsequent experiments, McCann revised the previous assumption and proposed that spatial structure measured in the illumination can affect the colors, leading to the departure from perfect color constancy [14, 15]. At the same time, many advances have been made in image enhancement methods based on this theory. Guo et al. [10] proposed an effective method called LIME, which refines the initial illumination map by using a structural prior as the final illumination map and further denoising using the method of block matching 3-D filtering algorithm (BM3D) [16], Wang et al. [17], and proposed a method called naturalness preserved enhancement algorithm (NPEA), which used a bi-log transformation to process the illumination. Kong et al. [18] introduced the Poisson function into the retinex model and proposed a Poisson noise aware retinex model (PNAR). Unfortunately, these methods still struggle to balance over- and under-enhancement in complex lighting situations and do not give a good sense of naturalness.

Fusion-based methods decompose the input image into multiple sub-images based on frequency, gray-scale or other criteria. Each sub-image will be processed separately and finally fused into the output image. These methods show excellent results in image enhancement for low-light images, especially those in complex lighting scenes. Ying et al. [14] proposed a bio-inspired multi-exposure fusion framework (BIMEF) which is a decomposition method based on the camera response model can effectively prevent the appearance of over-enhanced areas, but the contrast of the enhanced images is unsatisfactory. Hessel and Morel [19] used a multi-exposure model to enhance low-light images to varying degrees and then fuse the images. However, the loss of detail in the enhanced images is severe, and the color performance is not good enough. Peng et al. [20] performed adaptive gamma correction on multi-scale images separately and fused them, and Xu et al. [21] enhanced the original image with different non-linear mapping relationships and fused them. Both methods deal with individual regions of different brightness individually and do not take into account the dark–light relationship in the same image, so the enhanced image may appear to be brighter in areas with less real light than in areas with more light, resulting in the unnaturalness of the image. Wang et al. [22] proposed a method for edge-enhanced multi-exposure image fusion in YUV color space, with image decomposition, weight calculation and fusion based on Gaussian pyramids. This method achieves excellent results, but it is difficult to obtain a multi-exposure image of an identical scene in the real world, so the application of the method is limited. In the area of fusion-based image enhancement techniques, it has been observed that an increase in the number of artificial images potentially improves the outcome for low-light images existing within complex lighting environments. However, this augmentation invariably leads to a

proportional rise in computational time for the algorithm. In situations involving a singular target or those characterized by uniformly low illumination, the application of an excessive number of artificial images is rendered superfluous.

The deep convolutional neural network has developed rapidly and achieved many results in image processing area. Li et al. [23] proposed a convolutional neural network (CNN) image enhancement algorithm called Lighten-Net, which uses the CNN to estimate the illumination map, optimizing it with the guide filter and finally obtaining an enhanced image according to the retinex model. Zheng et al. [24] proposed a hybrid learning framework that compensates model-driven and data-driven methods with each other to generate multi-scale exposure images and fuse them. Wei et al. [25] proposed a deep network called Retinex-Net, including a network for decomposition and a network for illumination adjustment. Guo et al. [26] trained a lightweight deep network named DCE-Net, to estimate pixel-wise and high-order curves for dynamic range adjustment of a given image. Ma et al. [27] developed a new context-sensitive decomposition network (CSD-Net) architecture to exploit the scene-level contextual dependencies on spatial scales. The performance of these methods depends heavily on the quality and quantity of the dataset, and it will be more desirable to collect images with different illuminations from similar scenes, which greatly increases the difficulty of obtaining a training set. Also, since subjective visual perception is more important than single metrics such as brightness, contrast, and sharpening when assessing the quality of enhanced images, this also poses a great difficulty in training the network.

3 Proposed method

Our proposed method revolves around identifying the major grayscale distribution ranges in the original low-light image. Then, we utilize the IAP-OCTM to selectively enhance each major grayscale range, resulting in a corresponding artificial image sequence. Subsequently, we perform image fusion based on the "pixel healthiness" of each pixel in each artificial image. This fusion process ensures that the final enhanced image aligns with human visual perception, while simultaneously enhancing contrast. The overall framework of our proposed method is depicted in Fig. 1.

3.1 Multi-scale parameter determination

Compared to the overall enhancement method, the multi-scale fusion enhancement method can pay more attention to the image details during the enhancement process, and effectively prevent over-enhancement and under-enhancement at the same time. In the fusion-based method, determining the number of artificial images to be generated and the degree

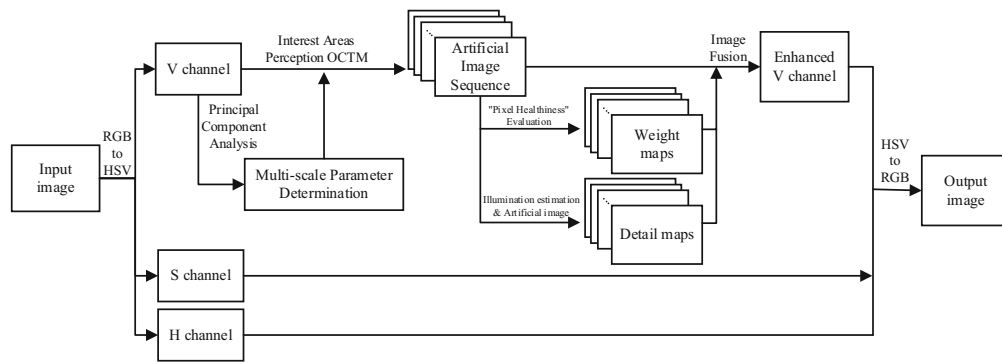


Fig. 1 Diagram of the proposed method

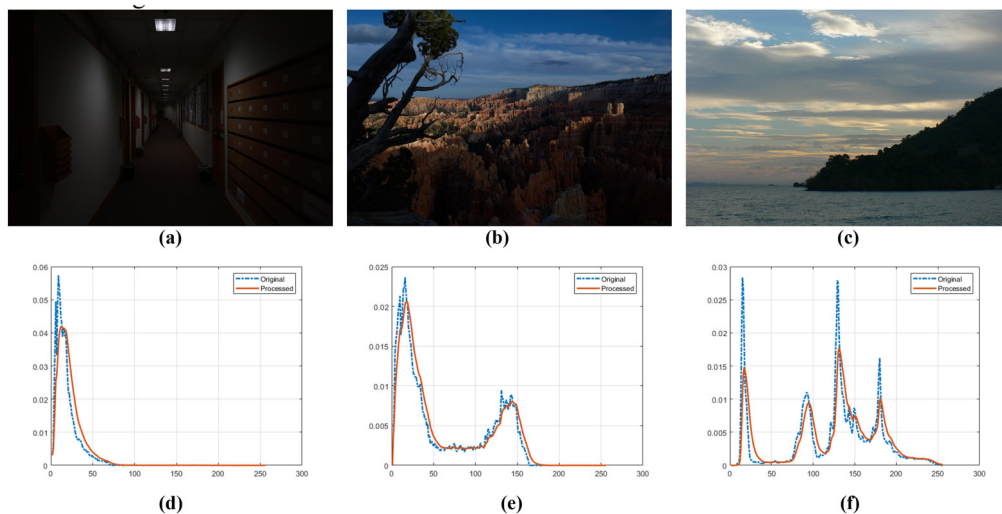


Fig. 2 Examples and V-channel histograms of three representative low-light images. **a–c** Low-light images. **d–f** Original histogram in blue, and processed histogram in orange

of enhancement of each image are one of the keys to determining the quality of the output image; in this paper, we call this step as multi-scale parameter determination. In order to find a method to adaptively determine the enhancement intervals for different kinds of low-light images, we analyzed histogram statistics from over 1000 low-light images from six datasets (VV-data [28], LIME-data [29], NPE-data [17], DICM [30], SCIE-data [6], and MEF-data [31]). We will illustrate the flow of our method specifically with several typical types of low-light images shown in Fig. 2.

For images directly captured by the camera, the camera position, target position and lighting conditions are fixed. So, for human visual system (HVS) observation, a main target in the picture, or the background part, will be shown as a piece of area with a similar gray value. In this paper, they are called "HVS interest areas," and these are the parts that need to be focused on in the process of image enhancement. The histogram h showing the grayscale distribution can accurately reflect these grayscale intervals that need to be focused on in

the enhancement process through the form of a peak. Therefore, we believe that based on the histogram distribution, we can better determine how many generated images are needed to enhance a low-light image, and determine which part of the grayscale intervals should be focused on in these generated images.

As shown in Fig. 2, in some indoor scenes, the illumination conditions are uniform, and the low-light images are overall in a similar grayscale interval, from the histogram, it is also obviously that there is only one peak, so we can degrade the fusion enhancement method in this paper to a single-image enhancement scheme. On the other hand, the natural outdoor scenes shown in Fig. 2b and c have widely varying grayscale values due to the presence of close-up and distant views, as well as different materials such as trees, water, and sky. For example, the trees and rocks parts in the shadows in Fig. 2b are clustered in the histogram to form a peak in the grayscale range of 0–50, while the sky in the distant view and the rocks in the light form another peak in the grayscale range of 110–160. Figure. 2c obviously shows

that the trees on the island are the darkest and they cluster in the grayscale range of 10–30, the water portion has a grayscale range of about 75–105, the large area of clouds in the sky mainly occupy the grayscale range of 115–160, and the remain sky is the brightest which has the grayscale value around 180. Thus, for Fig. 2b, a better solution would be to use two generated images one focusing on enhancing the close-up portion while letting the sky portion overexposed, and the other enhancing the distant portion while keeping the close-up views underexposed. The two images are later fused to obtain an enhanced image with all parts at the appropriate level of enhancement. For Fig. 2c, we will use four generated images to enhance the trees, the water, the clouds, and the sky, respectively, and then perform fusion.

Here, we will specifically describe the steps of determining the number of generated images n and finding the corresponding enhanced grayscale range boundaries th . First step in our method, in order to focus on the overall trend of the histogram, we processed the curve using Kalman filter to fit the overall trend while removing the small peaks. The filtered histogram is denoted by h_k . Since the requirement for the effect of the Kalman filter here is only to highlight certain trends, the choice of parameters is more flexible. In this paper, the process noise covariance Q is set to 0.0001, and the measurement noise covariance R is set to 0.01, which can get satisfactory results for all test images. Also, in order to correct for the overall image grayscale bias that exists in some low-light images, such as the image looking overall grayish, we add a parameter th_0 to mark the starting point of the gray value in current image, th_0 is set to 0.05% in the cumulative histogram.

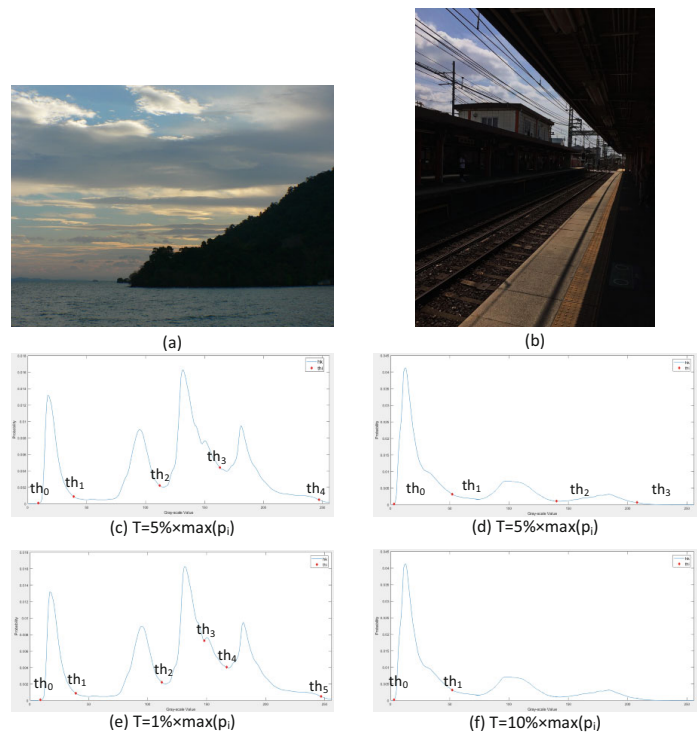
In the second step, we first identify all the peaks in the histogram. Although we use the Kalman filter to filter the histogram, it is still not possible to guarantee that the boundary determination will not be affected by small fluctuations in a complex variety of image types. Therefore, after finding the peaks of the histogram, we will calculate the height of prominences for each peak denoted as p_i . To further filter out ups and downs due to small grayscale variations in the same observation target, after running through all the peaks, we denote 5% of the maximum value of all the height of prominences as a judgmental threshold, $T = \max(p_i) \times 0.05$.

We discarded the peaks whose $p_i < T$ in third step, as they are insufficient to represent an independent observation

target. When the histogram is relatively homogeneous, the probabilities corresponding to each pixel are also relatively average, and since the total number of pixels in the pixel image is fixed, the maximum value and prominence heights of the individual principal component peaks will also be low. As a result, the threshold T we calculated will also be lower, which will appropriately protect the smaller peaks from being discarded and prevent the omission of the observed target. when there is a clear distribution concentration in the histogram, it represents that there is a very obvious observed target in the image, then the threshold T will be larger at this time, and some small histogram peaks due to texture or details will no longer be considered as separate targets, which is a very intuitive way. Since the 5% is chosen to filter out some too-small peaks, this value is not particularly strict either, but it should not be too large or too small. Figure 3 gives the cases of the possible over-attention to tiny ups and downs and the omission of principal component peaks that can occur when this value is set to 1% and 10%. Figure 3e shows that when $T = \max(p_i) \times 1\%$ the processing of the histogram of Fig. 3a appears to be overly focused on the small peaks, making an extra enhancement boundary th_3 , which may lead to redundant computations in the algorithm. When $T = \max(p_i) \times 10\%$, it can lead to missing the enhancement boundaries should be considered, th_2 and th_3 , as shown in Fig. 3f, which can affect the quality of the enhanced result. When the threshold is set to 5%, Fig. 3c and d shows that we can achieve the target quite well.

Next are the final two steps of our algorithm, we have determined the center gray value of the principal component peaks of an image. For each aggregated peak, 95% of the peak prominence is sufficient to encompass that principal component peak. After that, we look for the first gray value after the peak, which probability value less than 5% of the peak prominence height, as the boundary, denoted as $(th_1, th_2, \dots, th_n)$. Here, the total number of peaks n is the size of artificial generated images. And the grayscale range of interest for each generated image $I_i (i \in [1, 2, \dots, n+1])$ can then be represented as $th_0 \sim th_1, th_0 \sim th_2, \dots, th_0 \sim th_n$. The algorithmic process for multi-scale parameter determination is summarized in Algorithm 1.

Fig. 3 Example for enhanced grayscale boundaries determination with different T . **a** and **b** Low-light images. **c** and **d** Comparison of 5% and 1%. **e** and **f** Comparison of 5% and 10%



Algorithm 1: Multi-scale Parameter Determination

Input: Histogram of the image, Kalman filter parameters Q and R

Step 1: Kalman filter is performed on the histogram, and the result is h_k . Calculate the cumulative histogram, using the gray value corresponding to 0.5% as th_0 .

Step 2: Find all the peaks in h_k and calculate the height of the prominence of each peak, denoted as p_i , and find the peak filtering threshold $T = \max(p_i) \times 0.05$

Step 3: Remove small peaks, which $p_i < T$, and the number of remaining peaks n is the size of the generated images

Step 4: For each of the remaining peaks, from the histogram, look for the first gray level x after the peak that satisfies $h_k(x) < p_i \times 0.05$, then for this principal component peak, the gray-boundary of the gray-scale range of interest $th_i = x$.

Step 5: The gray-scale boundary of region of interest of the input low light image is obtained, denoted as $(th_1, th_2, \dots, th_n)$.

Output: the size of the generated images n , Boundary of the gray scale interval that each generated image wants to enhance $th_i, (i \in n)$

3.2 Interest-areas perception OCTM for artificial image sequence

In our method, the purpose of each artificial image is to enhance the specific grayscale interval to a suitable brightness. Therefore, global enhancement methods (e.g.,

histograms and some retinex-based methods) have difficulty in achieving this regional enhancement goal. Instead, we propose an interest-areas perception optimal contrast-tone mapping (IAP-OCTM) method based on general OCTM theory, which generates a set of suitable parameters for each generated image according to the multiscale parameters obtained previously. In the specific calculation process,

IAP-OCTM can find the most suitable gradient for each gray level, and the final mapping curve is the accumulation of gradients.

The increment of two adjacent gray levels in mapping curve, also be known as gradient, can represent the difference in gray values between those in the image after mapping. Thus, a larger gradient in the mapping curve at a particular gray level means this gray level has a greater contrast in the enhanced image. However, if the gradient is unconstrained, the way to maximize the global contrast is to binarize the image at the gray value with the highest probability, which is clearly unreasonable. Therefore, in general OCTM method [10], as shown in Eq. (1), there are upper and lower constraints on the gradient, thus making the process of solving the enhancement curve a linear optimization problem.

$$\begin{aligned} & \arg \max_s \sum_{j \in L_{in}} p_j s_j \\ & s.t. (a) \sum_{j \in L_{out}} s_j \leq L_{out} \\ & (b) s_j \geq \varepsilon_{min} \\ & (c) s_j \leq \varepsilon_{max} \end{aligned} \tag{1}$$

here p_j is the probability density of the gray level j in original low-light image, s_j is the gradient of the mapping curve at this gray level, and L_{in} and L_{out} are the dynamic range before and after the process, respectively. We need to ensure that the output image is still in the grayscale range of 0 to 255, so L_{out} need to be set to 0–255. ε_{min} and ε_{max} are the upper and lower limits of the s_j , respectively, and these three parameters will directly affect the enhancement process and are the variable parameters that need to be adaptively determined for different generated images for our proposed IAP-OCTM.

In IAP-OCTM, global enhancement is not required for each generated image, but focused on specific gray scale intervals. Therefore, fixed parameters are no longer applicable. We improve the parameter computation method in general OCTM for adaptive generating the constraints.

$$\begin{cases} L_{in}^i = th_0 \sim th_i \\ \varepsilon_{min} = \frac{1}{3} \\ \varepsilon_{max} = P^2 / L_{in}^i \end{cases} \tag{2}$$

Equation (2) gives the calculation of the three parameters in the IAP-OCTM. The first one is L_{in} . L_{in} determines which portion of the grayscale interval will be stretched to the interval from 0 to 255. In contrast to the simple way of setting L_{in} as 0 ~ 255 in OCTM, we set L_{in}^i for the image processing for I_i as $th_0 \sim th_i$. The second one is ε_{min} , determines the lower limit of the gradient, The smaller the ε_{min} , the better the tone continuity, but the degree of enhancement

will also be limited. In our experiments, a pleasing visual appearance is obtained when ε_{min} is set to $\frac{1}{3}$. The third one is ε_{max} , which is the upper limit of the gradient, prevents the gradient assignment from being too aggressive, resulting in too few gray levels in the enhanced image. ε_{max} is directly related to the target range of gray-scales. When the current image focuses on darker parts, the th_i will be small. At this point, we will require a stronger enhancement to stretch out the details hidden in dark areas. So, we should set a larger value for ε_{max} . If the current image intends to enhance the brighter regions of the input image, th_i is a larger value. In this case, the dark areas have been well-enhanced in another artificial image, so we do not need to be too aggressive with the enhancement but should pay more attention to preventing over-enhancement, and the ε_{max} should be smaller. We constructed a function to represent this inverse relationship mentioned above, as shown in Eq. (2). P is the parameter that adjusts the sensitivity of the inverse function, which varies from 18 to 22. In the absolute majority of low-light images, $P = 20$ gives satisfactory results.

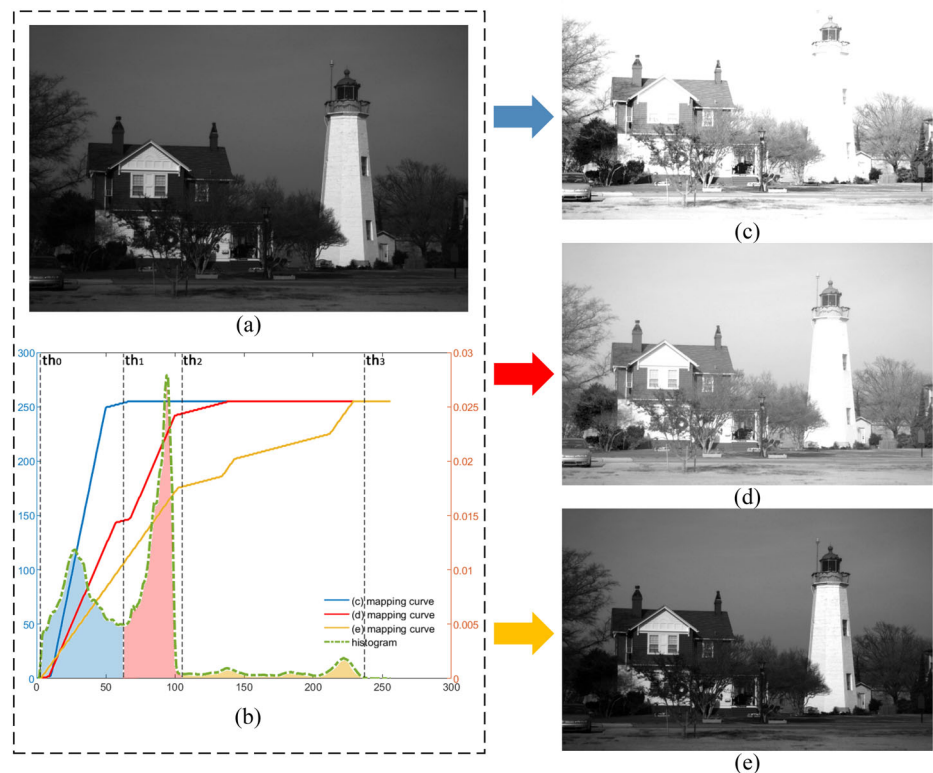
So far, the parameters in Eq. (1) and Eq. (2) are determined. We can solve the linear optimization equation to obtain the gradient s_j for each generated image. And we can easily get the ideal mapping curve $M(j)$ by accumulating the gradients s_j . The enhanced artificial image I can be achieved from the original low-light image V by grayscale mapping according to the mapping curve, as shown in Eq. (3).

$$\begin{aligned} M(j) &= s_1 + s_2 + s_3 + \dots + s_j \quad (j \in [0, 1, \dots, 255]) \\ I_i(x, y) &= M_i(V(x, y)) \quad (i \in [1, 2, \dots, n+1]) \end{aligned} \tag{3}$$

here (x, y) represents a specific pixel location, I_i represents the i^{th} artificial image, and M_i represents the i^{th} mapping curve.

Figure 4. shows the processing of an example low-light image by the IAP-OCTM. In order to verify the effectiveness of our method, we take the *lighthouse* (Fig. 4a) as an example. Based on the multi-scale parameter $[th_i = th_0, th_0, th_3 \quad n = 3]$ (shown in Fig. 4b), we divide the input image into three "HVS interest areas." The blue area is the darkest part and requires the strongest enhancement via the blue mapping curve. The red area is the next darkest part, occupying the largest proportion of the whole picture. This area will be enhanced through the red mapping curve. At the same time, the yellow area is the brighter part of the original image and does not require aggressive enhancement. We should focus on improving the contrast of the target in the image and is enhanced via the yellow mapping curve. We can happily find that the enhancement results are as expected, as shown in Fig. 4c–e. In Fig. 4c, the darkest areas are properly enhanced, such as the top of the lighthouse and the chairs inside the car. In Fig. 4d, the image’s overall brightness is enhanced, especially in the parts of the sky and grass that

Fig. 4 Generation process of multi-scale artificial image sequences. **a** Original V-channel. **b** V-channel histogram, target enhancement boundary and enhancement curve. **c–e** Generated artificial image sequence



occupy large areas. In Fig. 4e, the texture and contrast of the windows and lighthouse are protected and prevented from being over-enhanced.

3.3 Multi-scale image fusion based on “pixel healthiness” evaluation

Human visual perception of image quality is heavily influenced by brightness and contrast. Pixels with appropriate brightness and contrast are often considered “healthier.” Therefore, a comprehensive “pixel healthiness” evaluation method based on both global illumination and local contrast is proposed in this paper. In addition, based on the healthiness evaluation results, a function is constructed to calculate the weights of each pixel in the artificial image sequence for the final multi-scale image fusion. Meanwhile, we also enhance the details and textures to optimize the fused images further.

3.3.1 Global illumination weight map generation

In each artificial image, the pixels in the target grayscale range will be enhanced to the appropriate gray levels, while the rest may be over- and under-enhanced since they are not considered in the current image. Therefore, we need to label the pixels in the satisfying illumination range and give them higher weights in the fusion process. In this sector, the global illumination weight map will be determined by both

the current image’s target grayscale range and the pixel’s illuminance estimation.

For each pixel, the darker or brighter gray value due to local textures or details is not representative of the actual brightness of the current pixel. Therefore, the ideal illumination estimate should ensure it is as close as possible to the original image while retaining only meaningful and distinct structural boundaries and ignoring details and textures within the same structure. We construct an efficient linear optimization equation to accomplish this. The first term of the formula is used to represent the difference between the luminance estimation and the original image, while the second term reflects the smoothness of the texture. This means that we need the first term as the dominant component when the pixel is at a structured edge, thus preserving the information in the original image and increasing the weight of the second term to smooth the texture when the pixel contains only texture, as shown in Eq. (4).

$$\arg \min_T \sum_{(x,y)} \left((T(x,y) - V(x,y))^2 + \lambda \times \sum \frac{M_{\text{com}}(x,y) (\nabla_{\text{com}} T(x,y))^2}{|\nabla_{\text{com}} V(x,y)| + \varepsilon} \right) \quad (4)$$

here $V(x,y)$ is the gray value of a pixel located on (x,y) in the V channel of the HSV color space, as the initial estimation. $T(x,y)$ is the final illumination estimation. ε is a very

small constant to avoid the zero denominator. The gradient operator ∇_{com} and weight coefficient M_{com} are the key point in Eq. (4).

We know that if there is a major edge in a local window, it will have a large directional gradient compared to a local window with only complex textures. Based on this, we construct four directional gradient detection matrices for horizontal (f_h), vertical (f_v) and diagonal directions (f_{dl}, f_{dr}). The directional gradient operator ∇_{com} is defined as the sum of the absolute values of the convolution sum of the four detection matrices with the image window, as shown in Eq. (5). Here, $\text{win}(x, y)$ is a 3*3 window centered at current pixel.

$$\begin{aligned}
 f_h &= \begin{bmatrix} -1/3 & -1/3 & -1/3 \\ 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} & f_v &= \begin{bmatrix} -1/3 & 0 & 1/3 \\ -1/3 & 0 & 1/3 \\ -1/3 & 0 & 1/3 \end{bmatrix} \\
 f_{dl} &= \begin{bmatrix} -1/6 & -1/3 & 0 \\ -1/3 & 0 & 1/3 \\ 0 & 1/3 & 1/6 \end{bmatrix} & f_{dr} &= \begin{bmatrix} 0 & -1/3 & -1/6 \\ 1/3 & 0 & -1/3 \\ 1/6 & 1/3 & 0 \end{bmatrix}
 \end{aligned} \tag{5}$$

$$\begin{aligned}
 \nabla_{\text{com}}(x, y) &= |\text{win}(x, y) * f_h| + |\text{win}(x, y) * f_v| \\
 &+ |\text{win}(x, y) * f_{dl}| + |\text{win}(x, y) * f_{dr}|
 \end{aligned}$$

To identify the major edges as accurately as possible, we use the inverse of the sum of the directional gradients of all nine pixels in the 3×3 window in which the current pixel is located as the negative correlation coefficient M_{com} , as Eq. (6) shows.

$$M_{\text{com}}(x, y) = \frac{1}{\sum_{(l,n) \in \omega_{\text{in}}(x,y)} \nabla_{\text{com}} V(l, n) + \varepsilon} \tag{6}$$

here ε is a very small constant to avoid the zero denominator. (l, n) represents 9 pixels in a window.

After computing the illumination estimation $T(x, y)$, we also propose an exponential-like evaluation function, as shown in Eq. (7), which integrates the illuminance estimation and the target grayscale range for each image to obtain a weight map representing the "illuminance healthiness" of each pixel.

$$w_{\text{ge}}^i(x, y) = \exp\left(-\frac{|T_i(x, y) - \text{th}_i|}{\sigma_{\text{ge}}}\right) \tag{7}$$

here w_{ge}^i is the global illumination weight map for i^{th} artificial image. σ_{ge} is an experimental parameter, set to 0.4 in this paper, and th_i is the boundary for i^{th} artificial image's target grayscale range in multi-scale parameter.

As shown in the second column of Fig. 5a is mainly responsible for enhancing the darkest parts to the proper brightness, such as the shrubs at the bottom left. The global illumination map in Fig. 5d shows that the corresponding

area is indeed given the maximum weight as well. Figure 5b is mainly used to enhance the main areas of the image that are at a medium gray level, so Fig. 5e shows that areas such as grass and trees have larger weights. Moreover, Fig. 5c is used to protect the local bright parts of the image and enhance the contrast. Therefore, in Fig. 5f, only the lighthouse and window with higher brightness in the original image are given larger weights.

3.3.2 Local contrast weight map generation

Empirically, it is widely accepted that the human eye has the best discrimination in moderate gray levels. The luminance adaptation model based on the just-noticeable-difference (JND) model summarized by Chun et al.[32] also verified it. In this paper, we have constructed a low-light just-noticeable-difference (LLJND) model based on the JND model, combined with the characteristics of low-light images as follows:

$$\begin{aligned}
 \text{LLJND}(I(x, y)) &= \begin{cases} 15 \times (1 - \sqrt{\frac{I(x, y)}{I_{\text{comfort}}}}) + I_{\text{res}}, & I(x, y) \leq I_{\text{comfort}} \\ \frac{I_{\text{res}}}{I_{\text{comfort}} + 1} \times (I(x, y) - I_{\text{comfort}}) & \\ + I_{\text{res}}, & \text{otherwise} \end{cases}
 \end{aligned} \tag{8}$$

I_{comfort} is the most comfortable gray level for the human eye in low-light images. For natural low-light images, humans usually perceive that it should be slightly lower than the image in normal light, so we set I_{comfort} to 100 there. I_{res} is the extreme grayscale resolution of HVS in the most comfortable situation, set to 3 in this article, which means that the human eye can distinguish pixels with a gray value difference of 3.

Based on the LLJND model, we constructed the pixel local contrast evaluation method. Not only does it take into account the gradient sum of the window in which the current pixel is located, but it also considers the human ability to perceive grayscale differences at the current gray level. At this point, we can obtain a local contrast evaluation criterion more in line with human observation habits, as shown in Eq. (9)

$$C_{\text{pixel}}^i(x, y) = \sum_{(l,n) \in \text{win}(x,y)} \frac{|I(l, n) - I(x, y)|}{\text{LLJND}(T(x, y))} \tag{9}$$

here $C_{\text{pixel}}^i(x, y)$ represents the local contrast of pixel (x, y) in the i^{th} artificial image. $\text{win}(x, y)$ is a 3×3 window, and (l, n) represent the pixels in this window. $T(x, y)$ represents the illumination estimation for current pixel obtained before.

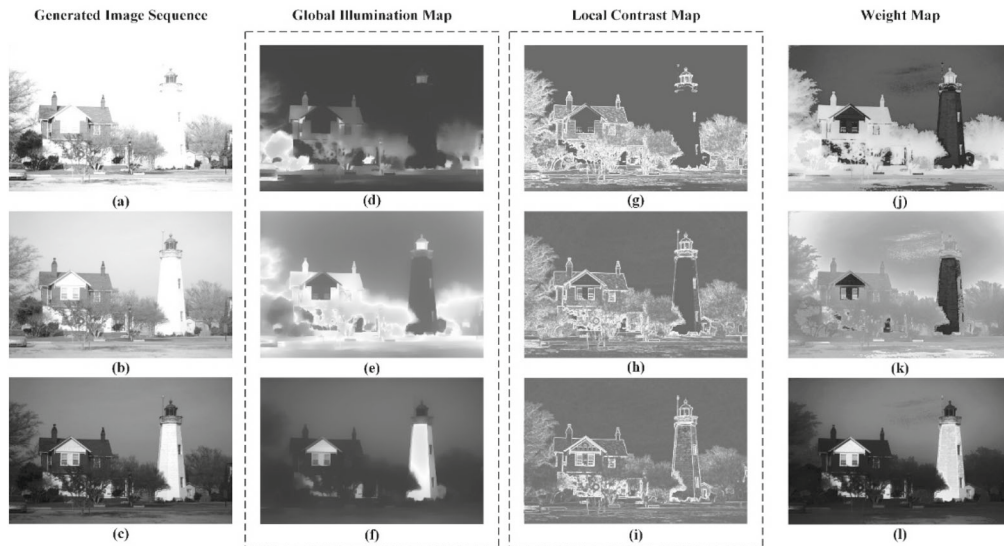


Fig. 5 Weight map sequence generation process. a–c Generated images sequence. d–f Global illumination map for generated image sequence. g–i Local contrast map for generated image sequence. j–l Final fusion weight map

Like the weight calculation function for global illumination, we also construct an evaluation function to assess the "contrast healthiness" based on the difference between the median contrast of the current image and the local contrast of the current pixel. As shown in Eq. (10), $w_{lc}^i(x, y)$ is the final obtained local contrast weight map for current artificial image. $\text{median}(\ast)$ is the operation to find the median and σ_{lc} is an empirical parameter to control the sensitivity, which is set to 0.4 here.

$$w_{lc}^i(x, y) = \frac{1}{1 + \exp\left(\frac{\text{median}(C_{\text{pixel}}^i(x, y)) - C_{\text{pixel}}^i(x, y)}{\sigma_{lc}}\right)} \quad (10)$$

As shown in the third column of Fig. 5, in local contrast maps of each artificial image (Fig. 5g–i), our evaluation method identifies the details and textures that show the best contrast in the current image and assigns higher weights to these pixels. Moreover, it is worth mentioning that the highlighted detail parts are consistent with the grayscale regions we assigned for the desired enhancement, proving that our generation for multi-scale images is as expected.

Finally, we define the final "pixel healthiness" based fusion weights $W^i(x, y)$ for the artificial image sequence as the normalized sum of the global illumination map and the local contrast map, as shown in Eq. (11).

$$W^i(x, y) = \frac{w_{ge}^i(x, y, i) + w_{lc}^i(x, y, i)}{\sum_{i=1}^{i=n+1} (i_{ge}(x, y, i) + w_{lc}^i(x, y, i))} \quad (11)$$

The fourth column of Fig. 5 is the calculated weight map. We can see that the grayscale range that each image wants to

enhance can also be well reflected in it. At the same time, the use of local contrast allows details, especially the edges of the regions, to be well protected compared to the global illumination map alone. This also proves that "healthier" pixels are indeed given higher fusion weights in our method. For example, shrubs and trees in Fig. 5j, sky and grass in Fig. 5k, and windows and lighthouse’s walls in Fig. 5l all have higher gray values in the weight map.

3.3.3 Detail and texture extraction

Sharper edges and textures can effectively enhance the overall image viewing experience, but valuable detail information is inevitably lost during the enhancement and fusion process. In order to further optimize the details and improve the quality of the fused image further, we also propose a fast detail enhancement method. As previously stated, the illumination estimation retains only the boundaries of areas with significantly different shades of gray, while smoothing out details and textures in the same structure. In our method, with the aim to extract the detail as fast and efficiently as possible without adding extra computational cost, we make full use of the illumination estimation obtained before, and the detail map can be very easily represented as the difference between the original artificial image and the illumination estimation, as shown in Eq. (12).

$$D^i(x, y) = I^i(x, y) - T^i(x, y) \quad (12)$$

here $D^i(x, y)$ represents the detail map, and $I^i(x, y)$ and $T^i(x, y)$ are the i th generated artificial image and the illumination estimation obtained before. It is essential to note that here $D^i(x, y)$ is allowed to be negative.

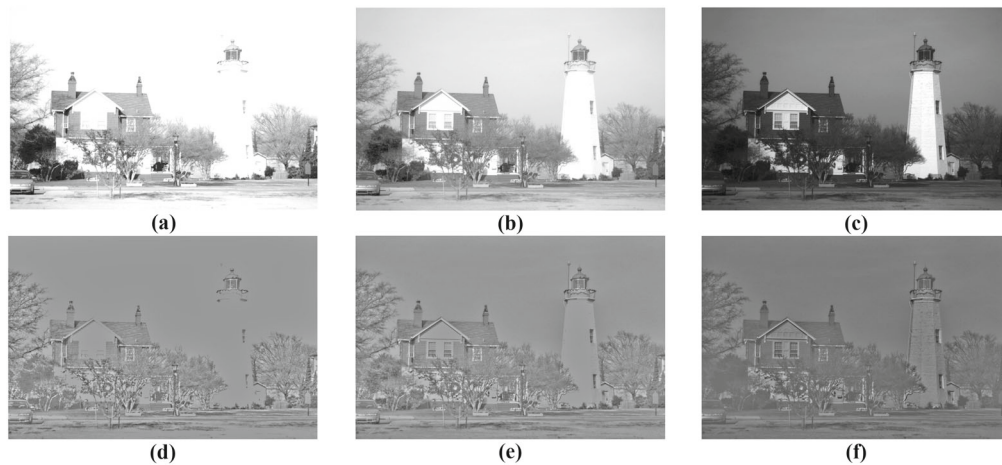


Fig. 6 Detail extraction effect diagram. **a–c** Artificial images sequence. **d–f** Detail extraction for each generated image

As shown in Fig. 6, each detail extraction map is well done in its respective enhanced grayscale area. Such as the wall texture of the lighthouse, as shown in Fig. 6f, will be well protected in the final fused image.

3.3.4 Artificial image fusion

Image fusion is the process of combining artificial image sequences into an enhanced image. We calculate the weighted sum of the artificial image sequence and the detail map-based weight map, as shown in Eq. (13).

$$E(x, y) = \sum_{i=1}^{i=n} \left(I^i(x, y) + D^i(x, y) \right) \times W^i(x, y) \quad (13)$$

here $E(x, y)$ represents the enhanced V channel. $I^i(x, y)$ is the i th artificial image, $D^i(x, y)$ is the detail map, and $W^i(x, y)$ is the weight map with index i in the sequence.

At this point, we get the enhanced grayscale image, as shown in Fig. 7c. After that, to get the enhanced color image while keeping the original hue, we use it as a new V channel and convert it back to RGB color space together with H and S channels. The final enhanced color image and details are shown in Fig. 7d. We are pleased to see that the whole image has been nicely enhanced. The cars in the darkness have been enhanced to a comfortable gray level for viewing, and the texture of the lighthouse is well-protected at the same time.

4 Experiments and discussion

This section will use experimental analysis to confirm the method's efficacy. We compared the proposed method with eight state-of-the-art methods, including the LIME + BM3D [29], ICIP2019 [33], LR3M [34], STAR [35], IESDD [36],

PNAR [18], EGFM [1], EJOM [37], Zero-DCE [26] and CSDGAN [27]. In addition, we tested these methods on 57 low-light images in six public datasets: VV-data, LIME-data, NPE-data, DICM, SCIE-data, and MEF-data. It is well known that the choice of parameters in a low-light image enhancement method is decisive for the effectiveness of the enhancement. Therefore, to make a relatively fair comparison, we use the default parameters of each method for all test images. For our method, all parameters will be set to the values described in the previous section.

We implement all experiments using MATLAB R2020b on a PC with an Intel(R) Core(TM) i5-11,400 CPU @2.60 GHz processor. (No GPU acceleration was used.)

4.1 Qualitatively comparison

Similar to our method, ICIP1029, EGFM and EJOM are based on image decomposition and fusion. While these methods do a good job of preventing over-enhancement, they do not perform well in contrast enhancement and color expression. As shown in Fig. 8b and c, ICIP1029 and EGFM enhance the darker parts of the image to some extent, but the image is still darker overall. EJOM is great in terms of overall brightness enhancement, as shown in Fig. 8d, but it still has problems with contrast, as if a layer of white fog is above the image, resulting in less vivid colors. Thanks to the multi-scale fusion of the principal components, our method achieves good results in terms of both overall brightness enhancement and contrast enhancement. Figure. 8e, clearly shows that the grass and trees (top line image) and the murals (bottom line image) all have better contrast and color expression in the images, which is more in line with human observation habits.

LIME, LR3M, STAR, IESDD and PNAR all belong to methods based on retinex theory. The enhanced image is obtained by processing the reflection layer obtained

Fig. 7 Image before and after processing by the proposed enhanced framework. **a** Original V channel. **b** Original color image. **c** Enhanced V channel. **d** Enhanced color image

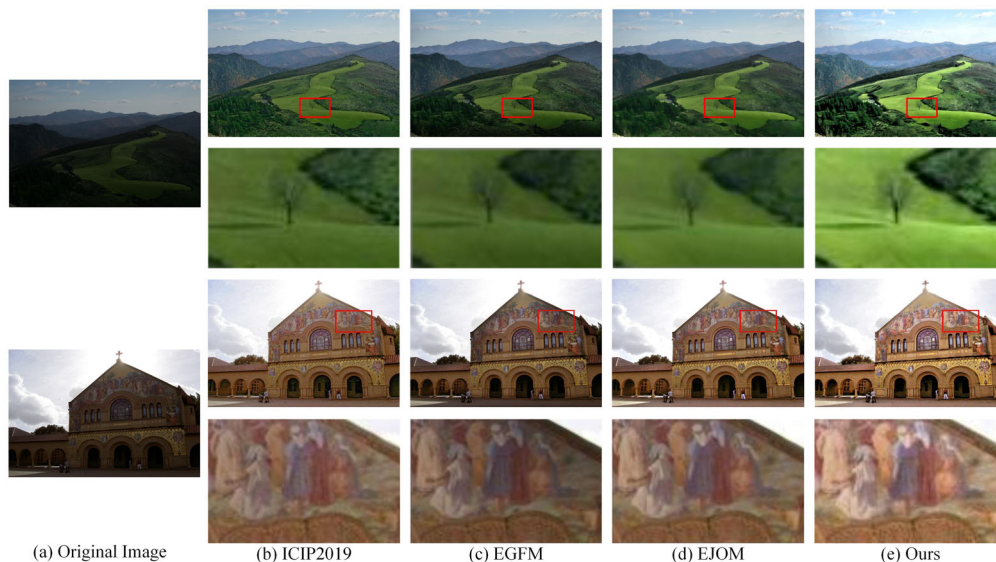
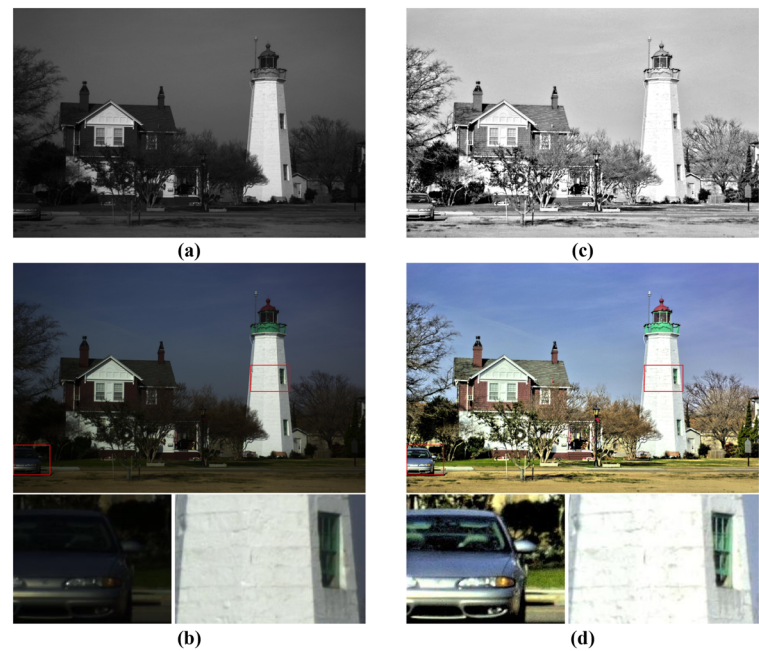


Fig. 8 Result comparisons between different methods. **a** Original image; **b** Results from ICIP2019; **c** Results from EGFM; **d** Results from EJOM; **e** Results from ours

from the decomposition. As shown in Fig. 9, all methods achieved acceptable results. LIME performed well in terms of overall brightness enhancement but also caused local over-enhancement leading to information loss and unsatisfactory edge processing. LR3M, STAR, IESDD, and PNAR showed excessive smoothing, resulting in a significant loss of image detail. As shown in Fig. 9c–f, the plant details (top line image), the forest texture (middle line image), and the signage text (bottom line image) all appear excessively blurred. IESDD is slightly better at detail protection, but the image is still relatively dark overall. By detailed comparison, the

results of our method are excellent. The brightness is sufficiently enhanced, and the detailed textures are well protected. The edges also show no anomalies that are uncomfortable to the human eye, as shown in Fig. 9g.

In addition, we compare our method with an open-source method based on deep learning (Zero-DCE and CSDGAN). Deep learning-based methods require a large amount of data for training and algorithms are complex. Especially when the resolution of low-light images is large, very large video memory of GPU is needed to support the algorithm to run,

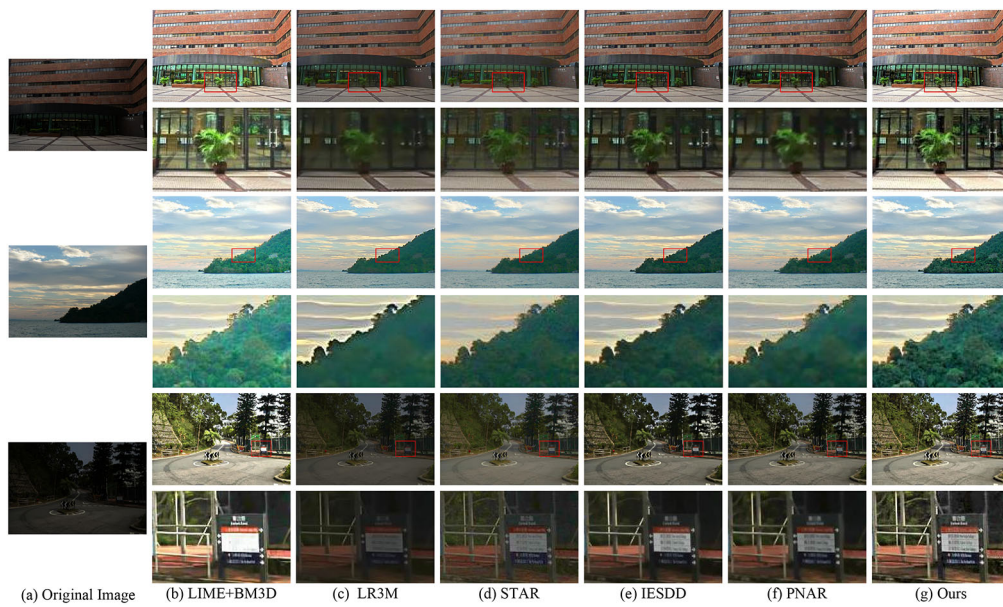


Fig. 9 Result comparisons between different methods. **a** Original image; **b** results from LIME + BM3D; **c** results from LR3M; **d** results from STAR; **e** results from IESDD; **f** results from PNAR; **g** results from ours

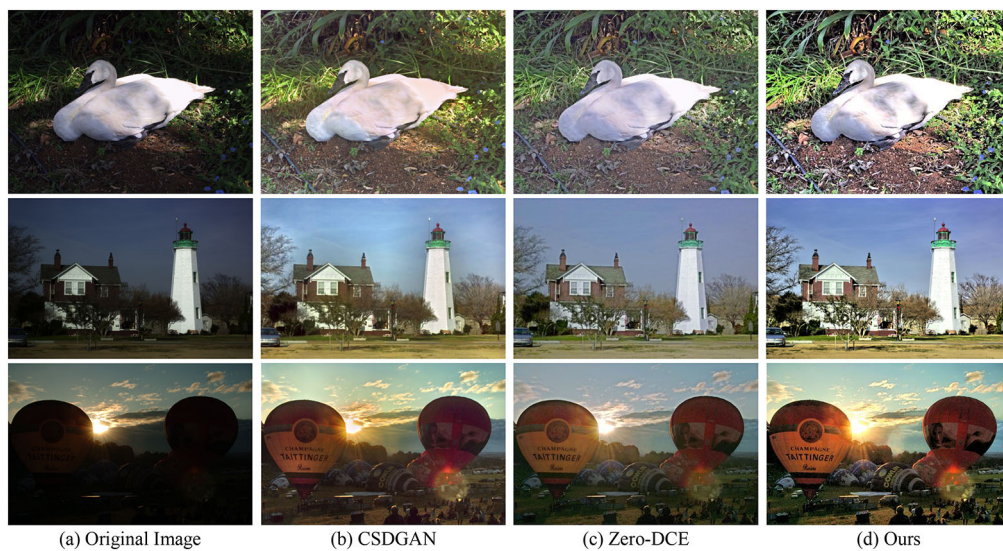


Fig. 10 Result comparisons between different methods. **a** Original image; **b** results from CSDGAN; **c** results from Zero-DCE; **d** results from ours

which brings a lot of difficulties for the use of the algorithm. We also found that the brightness of the CSDGAN and Zero-DCE enhanced images is appropriate, but the contrast performance is still unsatisfactory, and there is a layer of white fog in the vision. Meanwhile CSDGAN shows halo phenomenon in some low-light images, as shown in Fig. 10b. In contrast, our method maintains good enhancement effect and detail retention with better visual effect in processing all test images as shown in Fig. 10d.

In addition, we selected 57 images in the test dataset we used, with acceptable results for all 10 methods, to form a subjective evaluation dataset. Forty people engaged in image

processing research (divided into four groups randomly) conducted a personal evaluation of the enhancement results of all methods: 5 points for first place, 4 points for second place, 3 points for third place and zero points for the rest.

Table 1 summarizes the means of the four groups of scoring. We also calculated the variance between the four sets of scores and the percentage of scores for each method. We were pleased to find that our method stood out although CSDGAN, Zero-DCE and LIME + BM3D also achieved good scores. Moreover, the scores of our method showed good consistency between four groups. Table 1 demonstrates that the proposed

Table 1 Statistic results of subjective visual evaluation for 10 methods

Method	Scores (average of 4 group)	Variance (4 groups)	Percentage statistics (%)
ICIP2019	3.75	14.25	3.13
LIME + BM3D	12.25	10.92	10.21
LR3M	3.75	17.58	3.13
STAR	4.75	14.25	3.96
LESDD	3.5	32.33	2.92
PNAR	5.25	14.68	4.38
EGFM	6.5	15	5.42
EJOM	7	10	5.83
CSDGAN	23.25	16.92	19.38
Zero-DCE	17.5	11.67	14.58
Ours	33.75	8.92	27.08

method has a clear advantage regarding the intuitive perception of human eye observation.

4.2 Quantitative comparison

In order to make a quantitative comparison of image enhancement, we will evaluate the enhancement results of the 10 methods in terms of both image enhancement and image quality. Six widely recognized evaluation metrics are used.

4.2.1 Image enhancement

For image enhancement, we use three widely used metrics: absolute mean brightness error (AMBE) [38], discrete entropy (DE) [39] and measure of enhancement (EME) [40]. AMBE evaluates the luminance enhancement of the enhanced image compared to the original image. DE estimates the detail of the enhanced image based on a probability histogram distribution. EME evaluates the contrast of the enhanced result compared to the original image. Three metrics mentioned are where a higher score means better performance.

As shown in Table 2, our method achieved first place in DE and EME, showing that our results have the best performance in terms of detail and information protection. On AMBE, we achieved third, due to our method's commitment to finding the most appropriate enhancement strength rather than the stronger, the better. Both LIME + BM3D and CSDGAN, which achieved the top two places, are prone to local over-enhancement.

4.2.2 Image quality

For image quality, we adopt contrast enhancement based contrast-changed image quality measure (CEIQ) [41], naturalness image quality evaluator (NIQE) [42] and blind/referenceless image spatial quality evaluator (BRISQUE) [43] for the assessment of enhanced image quality. CEIQ considers structural similarity, histogram entropy, cross-entropy and other factors to assess image quality. NIQE characterizes the naturalness of an image by calculating the natural scene statistics (NSS) features of the test image patches. BRISQUE uses scene statistics with locally normalized luminance coefficients to quantify an image's possible loss of naturalness. The higher the CEIQ metric, the more information the image has and the better the overall image quality.

As shown in Table 3, our method ranks two firsts and one second among the three metrics characterizing overall image quality. The superiority of our method in the CEIQ metric is due to the full use of the entire gray-scale interval in our enhancement method, ensuring that the enhanced image has the highest possible information content and preventing over- and under-enhancement wherever possible. The lead in the NIQE and BRISQUE metrics is due to our principal component-based analysis and a fusion method that fully follows the viewing habits of the human-eye. This allows our method to produce enhanced images with the best human-eye viewing experience. This is why the enhanced image obtained by our method won by a large margin in the subjective test.

4.3 Ablative analysis

In this section, ablative experiments and analysis were performed to verify the effectiveness of each main component. Firstly, we discuss the necessity for multi-scale parameter determination in the Interest-areas perception OCTM by visual comparison and metrics of DE, EME, and CEIQ. In interest-areas perception OCTM, the threshold is determined based on the number of principal components in the image statistic, which we replace with a simple use of middle grayscale as the demarcation (image is divided into [0,127] and [128,255]) and experiment with images in the dataset. Since multi-scale parameter determination focuses on the adequate enhancement of individual regions in an image and prevents over-enhancement, it leads to better overall contrast and protection of detailed information. We provide the objective evaluation results for DE, EME and CEIQ for the images in the dataset in Table 4. We are pleased that multi-scale parameter determination performs better in all three metrics. Figure 11 clearly illustrates that our method has a clear advantage in the enhancement of the indoor part of the image.

Table 2 Comparison of average score of AMBE, DE and EME for 10 methods on six datasets

Methods	ICIP 2019	LIME + BM3D	LR3M	STAR	LESDD	PNAR	EGFM	EJOM	CSDGAN	Zero-DCE	Ours
AMBE↑	37.74	79.53	43.00	42.33	41.21	40.94	40.15	53.95	<u>65.67</u>	42.28	58.56
DE↑	7.11	<u>7.55</u>	6.94	6.99	7.19	7.20	7.27	7.36	7.31	7.20	7.61
EME↑	<u>9.62</u>	8.56	6.63	7.31	6.85	7.65	9.39	7.07	7.45	9.42	10.02

The first and second results are highlighted in bold and underline

Table 3 Comparison of average score of CEIQ, NIQE and BRISQUE for 10 methods on six datasets

Methods	ICIP 2019	LIME + BM3D	LR3M	STAR	LESDD	PNAR	EGFM	EJOM	CSDGAN	Zero-DCE	Ours
CEIQ↑	3.18	<u>3.47</u>	3.13	3.12	3.26	3.27	3.18	3.28	3.26	3.23	3.52
NIQE↓	<u>2.69</u>	2.87	3.40	2.66	2.97	2.95	2.70	2.93	3.17	3.11	2.52
BRISQUE↓	28.47	28.57	33.95	26.62	31.85	32.51	26.51	31.03	20.76	28.63	<u>26.26</u>

The first and second results are highlighted in bold and underline

Table 4 Comparison of average score of DE, EME and CEIQ for method with middle grayscale demarcation and method with multi-scale parameter determination

Methods	Method with middle grayscale demarcation	Method with multi-scale parameter determination
DE↑	7.28	7.61
EME↑	9.16	10.02
CEIQ↑	3.32	3.52

The first results are highlighted in bold

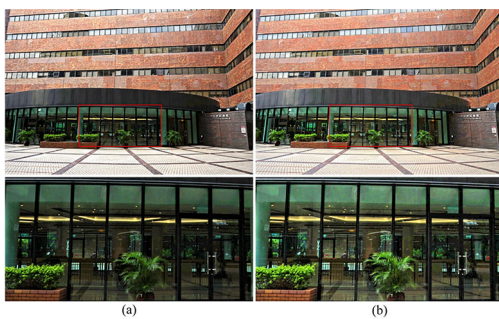


Fig. 11 Result comparisons between different methods. **a** Result from method with middle grayscale demarcation, **b** result from method with multi-scale parameter determination

Secondly, we have verified the necessity for two factors in the "pixel healthiness" evaluation by DE, EME, BRISQUE, NIQE and CEIQ. As described in Sect. 2.3, global illumination and local contrast characterize whether the current pixel is in a good state regarding luminance and contrast, thus guiding the generation of fusion weights. Therefore, both maps can achieve good image fusion and complement each other

to achieve better fusion results. We give examples of image enhancement using global illumination map only, using local contrast map only and both together, as shown in Fig. 12. We also present the average results of the objective evaluation metrics for the three methods in Table 5.

As shown in Fig. 12, when we only using the global illumination map for fusion, some areas are under-enhanced as the effects of internal detail and texture enrichment are not considered. When only using the local contrast map, the high-frequency areas may lead to halo artifacts at the edges, whereas when the two maps are used together, a better enhancement is achieved.

Finally, we verified the necessity for Detail and Texture Extraction and enhancement. Fig. 13 shows an example of image enhancement with and without detail and texture extraction. The average results for DE, EME, BRISQUE, NIQE and CEIQ are presented in Table 5. The enhanced images using detail and texture extraction are significantly better than the images without it, in both direct human eye observation and objective metrics. This is because the detail and texture extraction makes it possible to increase the gray level difference in a limited gray level range, thus improving contrast. This process is similar to the optimization of the combination of the black-level correction (BLC) and edge enhancement (EE) processes in the image signal processing (ISP) pipeline.

4.4 Comparison of computational costs

Table 6 demonstrates the average computational runtimes for enhancing ten low-light images with a resolution of 1280 * 960, across all methods considered for comparison. Since our method may choose different values of n for different images,

Fig. 12 Result comparisons between different methods. **a** Result from method with global illumination map only, **b** result from method local contrast map only, **c** result from our method

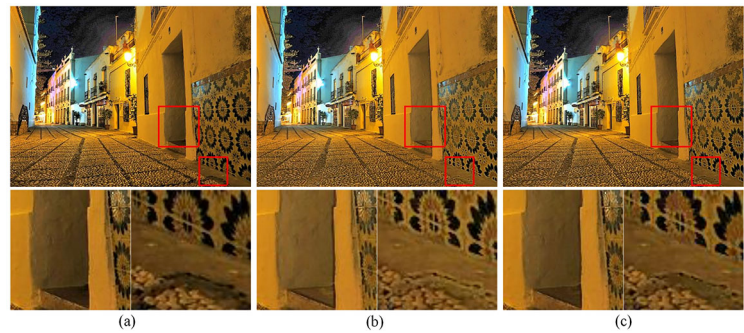


Table 5 Comparison of average score of DE, EME, BRISQUE, NIQE and CEIQ for method with global illumination map only, method with local contrast map only, method without detail and texture extraction and our method

Methods	Global illumination map only	Local contrast map only	Method without detail and texture extraction	Our method
DE \uparrow	7.32	7.33	7.23	7.61
EME \uparrow	9.20	9.25	7.84	10.02
BRISQUE \downarrow	26.33	28.15	28.35	26.26
NIQE \downarrow	2.94	3.02	2.95	2.52
CEIQ \uparrow	3.32	3.30	3.29	3.52

The frist results are highlighted in bold

we detail the runtime for n values ranging from 1 to 8. This delineation ensures a comprehensive and precise depiction of our algorithm's computational efficiency when applied to low-light images of varying complexity. In the comparison, we exclude the consideration of GPU acceleration for methods such as CSDGAN and Zero-DCE, as their performance is significantly influenced by such hardware acceleration. Relative to EGFM and EJOM, which also employ image fusion from multiple artificially generated images for enhancement, our methodology demonstrates comparable time consumption at $n = 2$ and $n = 3$. Although it does not surpass the rapidity of straightforward and efficient single-frame enhancement techniques like LIME and LESDD, it notably outperforms more computationally intensive single-frame enhancement strategies, including LR3M and PNAR.

The efficiency of our method is influenced by the value of n . Under complex lighting and scene conditions, a larger n value is necessary to ensure an optimal enhancement result, though this increases computation time. It is crucial to forecast computation times for different n values, necessitating a thorough analysis for each step, as shown in Table 7 and Fig. 14. This analysis combines a review of computation principles with experience data to estimate the computation time for each step accurately.



Fig. 13 Result comparisons between different methods. **a** result from method without detail and texture extraction, **b** result from our method

Table 6 Average computation time of different method (second)

Methods	Average computation time
ICIP2019	10.93
LIME + BM3D	<u>4.86</u>
LR3M	205.5
STAR	53.61
LESDD	3.12
PNAR	58.26
EGFM	18.974
EJOM	16.735
Our Method($n = 1$)	6.6696
Our Method($n = 2$)	11.4370
Our Method($n = 3$)	16.1928
Our Method($n = 4$)	20.8286
Our Method($n = 5$)	27.3887
Our Method($n = 6$)	33.4356
Our Method($n = 7$)	38.0765
Our Method($n = 8$)	43.1314

The frist and second results are highlighted in bold and underline

The process of multi-scale parameter determination is independent of the value of n , requiring only a single pass through the histogram statistics of the image. As depicted in Table 7, this phase consistently necessitates approximately 0.61 s across all n values, denoted as $t_{\text{step1}} = 0.61s$. In the step of artificial images generation, the IAP-OCTM will perform a uniform process of linear optimization calculation

Table 7 Average computation time for different step (second)

Step	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$
Multi-scale Parameter Determination	0.6133	0.6091	0.6102	0.6099	0.6135	0.6328	0.6198	0.6192
Artificial Image Generation	0.1422	0.2902	0.4398	0.5991	0.7228	0.8995	1.0331	1.1656
Global Illumination Weight Map Generation	2.359	4.1808	6.108	7.7601	10.59	12.9994	14.8068	16.7502
Local Contrast Weight Map Generation	3.4926	6.2843	8.9546	11.7695	15.3591	18.7861	21.4891	24.4633
Detail and Texture Extraction	0.0016	0.0029	0.0041	0.0052	0.0078	0.0087	0.0108	0.013
Artificial Image Fusion	0.0609	0.0697	0.0761	0.0848	0.0955	0.1091	0.1169	0.1201
Total Time	6.6696	11.437	16.1928	20.8286	27.3887	33.4356	38.0765	43.1314

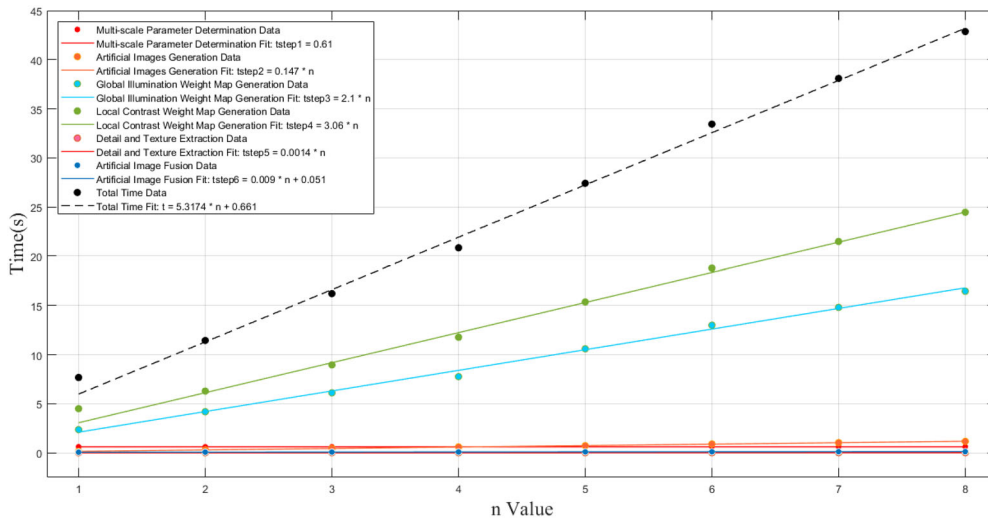


Fig. 14 Relationship of computation time for each step and the value of n

and grayscale mapping for each generated image. Hence, the time required for each synthetic image is constant, approximately 0.147 s for an image with a resolution of 1280*960. Therefore, the computational time for this step is directly proportional to the value of n , expressed as $t_{step2} = n \times 0.147s$. The generation of global illumination weight maps and local contrast weight maps represent the most time-consuming components within our method because both processes need iterating over every pixel of each artificial image and performing operations such as sorting, and convolution. The operations performed on each artificial image are the same thus, the time required also follows a proportional relationship. Specifically, the Global illumination weight map generation step requires approximately 2.1 s for each artificial image, and the local contrast weight map generation needs four convolution operations for per pixel, leading to an approximate duration of 3.06 s for each generated image. The computation time can be expressed as $t_{step3} = n \times 2.1s$ and $t_{step4} = n \times 3.06s$. The detail and texture extraction step involves a single subtraction operation, resulting in a significantly short processing time. This operation is consistently identical for every image, with computation time

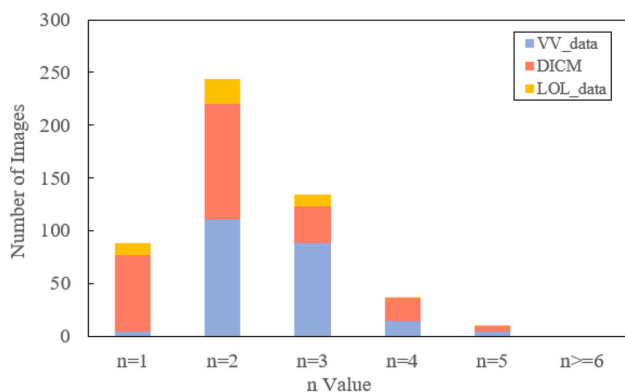
approximately being 0.0014 s hence, the duration of this step can be denoted as $t_{step5} = n \times 0.0014s$. The final procedure in the proposed method is artificial image fusion, which performs multiplication and addition operations n times, depending on the n value. Consequently, the time required for this step adheres closely to a linear function with n value. Through experimental data and fitting, we can represent this as $t_{step6} = n \times 0.009 + 0.661s$. Therefore, the correlation between the value of n and the algorithm's computation time can be represented as the summation of the aforementioned steps, denoted by Eq. (14).

$$t = t_{step1} + t_{step2} + t_{step3} + t_{step4} + t_{step5} + t_{step6} = n \times 5.3174 + 0.661s \tag{14}$$

Figure 14 vividly illustrates the comparison between experimental data and the fitting curve for varying n values. This relationship clearly demonstrates how the computation time escalates with increased n values. At this time, we can accurately predict the computation time of the proposed

Table 8 Statistical results of n value for images in three dataset and average computation time

Dataset	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n > = 6$	Average computation time (s)
VV_data	4	111	88	15	4	0	14.2132
DICM	73	109	35	21	5	0	11.8461
LOL_data	11	24	11	1	1	0	11.9835
sum	88	244	134	37	10	0	12.6809

**Fig. 15** Relationship of computation time for each step and the value of n

method for all n values, as it follows a simple linear relationship. This is crucial for optimizing the performance of our method under various complex lighting and scene conditions.

Since the n value has a decisive impact on the computational efficiency of the proposed method, and the images used in the algorithm description phase come from several datasets widely used in the field, it is necessary to further verify the efficiency of the proposed algorithm when processing typical low-light images captured in real-world scenarios. Therefore, we specifically evaluated the calculation results of the n value for each image in three publicly available low-light image datasets taken by ordinary users in practical scenes. It can demonstrate the efficiency of the proposed method in handling low-light images in real-world conditions more objectively.

The dataset include the VV_data [28] provided by Vasiliou Vonikakis, consisting of 222 of the most challenging images for enhancement, each containing parts that are correctly exposed and parts that are severely underexposed or overexposed. Additionally, the DICM dataset [30] includes 47 low-light images of everyday scenes captured with commercial digital cameras by Ying et al. Furthermore, the LOL_data [25], specifically collected for low-light enhancement by Wei, includes 244 images from real scenes. In total, these datasets contain 513 low-light images. Similarly, we scaled all images in the datasets to a resolution of 1280*960 to facilitate a clearer comparison of computational efficiency.

Table 8 and Fig. 15 present the statistical results of the n

value for 513 images. It is evident that for almost all real-world low-light images, the optimal n value computed by the proposed method is less than 5, and over 90% of the images in dataset having an n value less than 3. This indicates that the proposed method's computation time in practical applications is not a cause for concern. Additionally, Table 8 provides the average computation time for low-light images in each of the three datasets, as well as the average computation time for all 513 images, which is approximately 12.68 s.

Figure 16 also gives the enhancement effects of all eleven algorithms on the low-light images used in the Experiments and Discussion section.

5 Conclusion

In this paper, we propose a single-frame low-light image enhancement method based on multi-scale interest-area perception OCTM and "pixel healthiness" evaluation. In our method, the multi-scale parameter is determined adaptively by principal component analysis of the V-channel histogram. Based on this parameter, the interest-area perception OCTM generates artificial image sequences well according to the main target grayscale range in the image. Moreover, we use a "pixel healthiness" evaluation method based on a global illumination map and local contrast map to quickly and efficiently compute image fusion weights. In addition, we specifically protect and enhance edges and details based on illumination estimation. Subjective evaluation and objective metrics show that our algorithm performs better than existing single-frame image algorithms and other fusion-based algorithms in enhancement, contrast, color expression and detail retention. Meanwhile, the computation time of the method proposed in this paper is still long, especially when n is large, and is not advantageous compared with other single frame-based image enhancement methods. Our future work will focus on optimizing the global illumination weight map and local contrast weight map generation process. Moreover, enhanced image denoising is another future direction.



Fig. 16 Result comparisons between different methods

Acknowledgements All authors thank the National Natural Science Foundation of China, the National Defense Basic Research Program of China for help identifying collaborators for this work.

Funding National Funding for this study was obtained from the National Natural Science Foundation of China (No. U2141239).

Data availability The databases used in this paper are all publicly available datasets and can be accessed according to the references indicated. The code underlying the results presented in this paper is not publicly available at this time but may be obtained from the authors upon reasonable request.

References

- Lang, Y.-Z., Qian, Y.-S., Kong, X.-Y., Zhang, J.-Z., Wang, Y.-L., Cao, Y.: Effective enhancement method of low-light-level images based on the guided filter and multi-scale fusion. *J. Opt. Soc. Am. A* **40**, 1 (2023). <https://doi.org/10.1364/JOSAA.468876>
- Paul, A.: Adaptive tri-plateau limit tri-histogram equalization algorithm for digital image enhancement. *Vis. Comput.* **39**, 297–318 (2023). <https://doi.org/10.1007/s00371-021-02330-z>
- Srinivas, K., Bhandari, A.K., Singh, A.: Exposure-based energy curve equalization for enhancement of contrast distorted images. *IEEE Trans. Circuits Syst. Video Technol.* **30**, 4663–4675 (2020). <https://doi.org/10.1109/TCSVT.2019.2960861>
- Zhou, R., Wang, R., Wang, Y., Zhou, F., Guo, N.: Research on low-light image enhancement based on MER-Retinex algorithm. *SIVIP* **18**, 803–811 (2024). <https://doi.org/10.1007/s11760-023-02801-x>
- Ancuti, C.O., Ancuti, C., De Vleeschouwer, C.: Color balance and fusion for underwater image enhancement. *IEEE Trans. on Image Process.* **27**, 379–393 (2018). <https://doi.org/10.1109/TIP.2017.2759252>
- Cai, J., Gu, S., Zhang, L.: Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Trans. on Image Process.* **27**, 2049–2062 (2018). <https://doi.org/10.1109/TIP.2018.2794218>
- Zhang, H., Zhao, Q., Li, L., Li, Y., You, Y.: Multi-scale image enhancement based on properties of human visual system. In: 2011 4th International Congress on Image and Signal Processing. pp. 704–708. IEEE, Shanghai, China (2011)
- Zuiderveld, K.: Contrast Limited Adaptive Histogram Equalization. In: *Graphics gems*, pp. 474–485. Elsevier, Amsterdam (1994)
- Huang, S.-C., Cheng, F.-C., Chiu, Y.-S.: Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Trans. on Image Process.* **22**, 1032–1041 (2013). <https://doi.org/10.1109/TIP.2012.2226047>
- Xiaolin, Wu.: A linear programming approach for optimal contrast-tone mapping. *IEEE Trans. on Image Process.* **20**, 1262–1272 (2011). <https://doi.org/10.1109/TIP.2010.2092438>
- Ke, Gu., Zhai, G., Yang, X., Zhang, W., Chen, C.W.: Automatic contrast enhancement technology with saliency preservation. *IEEE Trans. Circuits Syst. Video Technol.* **25**, 1480–1494 (2015). <https://doi.org/10.1109/TCSVT.2014.2372392>
- Su, H., Yu, L., Jung, C.: Joint contrast enhancement and noise reduction of low light images via JND transform. *IEEE Trans. Multimed.* **24**, 17–32 (2022). <https://doi.org/10.1109/TMM.2020.3043106>
- Land, E.H.: The retinex theory of color vision. *Sci. Am.* **237**, 108–128 (1977). <https://doi.org/10.1038/scientificamerican1277-108>
- McCann, J.J., Parraman, C., Rizzi, A.: Reflectance, illumination, and appearance in color constancy. *Front. Psychol.* (2014). <https://doi.org/10.3389/fpsyg.2014.00005>
- McCann, J.J.: Retinex at 50: color theory and spatial algorithms, a review. *J. Electron. Imaging* **26**, 031204 (2017). <https://doi.org/10.1117/1.JEI.26.3.031204>
- Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. on Image Process.* **16**, 2080–2095 (2007). <https://doi.org/10.1109/TIP.2007.901238>
- Wang, S., Zheng, J., Hu, H.-M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. on Image Process.* **22**, 3538–3548 (2013). <https://doi.org/10.1109/TIP.2013.2261309>
- Kong, X.-Y., Liu, L., Qian, Y.-S.: Low-light image enhancement via poisson noise aware retinex model. *IEEE Signal Process. Lett.* **28**, 1540–1544 (2021). <https://doi.org/10.1109/LSP.2021.3096160>

19. Hessel, C., Morel, J.-M.: An extended exposure fusion and its application to single image contrast enhancement. In: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 137–146. IEEE, Snowmass Village, CO, USA (2020)
20. Peng, Y.-T., Liao, H.-H., Chen, C.-F.: Two-exposure image fusion based on optimized adaptive gamma correction. *Sensors*. **22**, 24 (2021). <https://doi.org/10.3390/s22010024>
21. Xu, Y., Yang, C., Sun, B., Yan, X., Chen, M.: A novel multi-scale fusion framework for detail-preserving low-light image enhancement. *Inf. Sci.* **548**, 378–397 (2021). <https://doi.org/10.1016/j.ins.2020.09.066>
22. Wang, Q., Chen, W., Wu, X., Li, Z.: Detail-enhanced multi-scale exposure fusion in YUV color space. *IEEE Trans. Circuits Syst. Video Technol.* **30**, 2418–2429 (2020). <https://doi.org/10.1109/TCSVT.2019.2919310>
23. Li, C., Guo, J., Porikli, F., Pang, Y.: LightNet: a convolutional neural network for weakly illuminated image enhancement. *Pattern Recogn. Lett.* **104**, 15–22 (2018). <https://doi.org/10.1016/j.patrec.2018.01.010>
24. Zheng, C., Li, Z., Yang, Y., Wu, S.: Single image brightening via multi-scale exposure fusion with hybrid learning. *IEEE Trans. Circuits Syst. Video Technol.* **31**, 1425–1435 (2021). <https://doi.org/10.1109/TCSVT.2020.3009235>
25. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement, <http://arxiv.org/abs/1808.04560>, (2018)
26. Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1777–1786. IEEE, Seattle, WA, USA (2020)
27. Ma, L., Liu, R., Zhang, J., Fan, X., Luo, Z.: learning deep context-sensitive decomposition for low-light image enhancement. *IEEE Trans. Neural Netw. Learn. Syst.* **33**, 5666–5680 (2022). <https://doi.org/10.1109/TNNLS.2021.3071245>
28. Vonikakis, V., Andreadis, I., Gasteratos, A.: Fast centre-surround contrast modification. *IET Image Process.* **2**, 19 (2008). <https://doi.org/10.1049/iet-ipr:20070012>
29. Guo, X., Li, Y., Ling, H.: LIME: low-light image enhancement via illumination map estimation. *IEEE Trans. on Image Process.* **26**, 982–993 (2017). <https://doi.org/10.1109/TIP.2016.2639450>
30. Ying, Z., Li, G., Gao, W.: A bio-inspired multi-exposure fusion framework for low-light image enhancement, <http://arxiv.org/abs/1711.00591>, (2017)
31. Ma, K., Zeng, K., Wang, Z.: Perceptual quality assessment for multi-exposure image fusion. *IEEE Trans. on Image Process.* **24**, 3345–3356 (2015). <https://doi.org/10.1109/TIP.2015.2442920>
32. Chou, C.-H., Li, Y.-C.: A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile. *IEEE Trans. Circuits Syst. Video Technol.* **5**, 467–476 (1995). <https://doi.org/10.1109/76.475889>
33. Ghosh, S., Chaudhury, K. N.: Fast bright-pass bilateral filtering for low-light enhancement. In: 2019 IEEE International Conference on Image Processing (ICIP). pp. 205–209. IEEE, Taipei, Taiwan (2019)
34. Ren, X., Yang, W., Cheng, W.-H., Liu, J.: LR3M: robust low-light enhancement via low-rank regularized retinex model. *IEEE Trans. on Image Process.* **29**, 5862–5876 (2020). <https://doi.org/10.1109/TIP.2020.2984098>
35. Xu, J., Hou, Y., Ren, D., Liu, L., Zhu, F., Yu, M., Wang, H., Shao, L.: STAR: a structure and texture aware retinex model. *IEEE Trans. on Image Process.* **29**, 5022–5037 (2020). <https://doi.org/10.1109/TIP.2020.2974060>
36. Hao, S., Han, X., Guo, Y., Xu, X., Wang, M.: Low-light image enhancement with semi-decoupled decomposition. *IEEE Trans. Multimed.* **22**, 3025–3038 (2020). <https://doi.org/10.1109/TMM.2020.2969790>
37. Lang, Y.-Z., Wang, Y.-L., Qian, Y.-S., Kong, X.-Y., Cao, Y.: Effective method for low-light image enhancement based on the JND and OCTM models. *Opt. Express* **31**, 14008 (2023). <https://doi.org/10.1364/OE.485672>
38. Chen, S.-D., Ramli, A.R.: Minimum mean brightness error bi-histogram equalization in contrast enhancement. *IEEE Trans. Consum. Electron.* **49**, 1310–1319 (2003). <https://doi.org/10.1109/TCE.2003.1261234>
39. Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423 (1948)
40. Agaian, S.S., Panetta, K.: Transform-based image enhancement algorithms with performance measure. *IEEE Trans. Image Process.* **10**, 16 (2001)
41. Yan, J., Li, J., Fu, X. (2019): No-reference quality assessment of contrast-distorted images using contrast enhancement. 15. <https://doi.org/10.48550/arXiv.1904.08879>
42. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a “completely blind” image quality analyzer. *IEEE Signal Process. Lett.* **20**, 209–212 (2013). <https://doi.org/10.1109/LSP.2012.2227726>
43. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. *IEEE Trans. on Image Process.* **21**, 4695–4708 (2012). <https://doi.org/10.1109/TIP.2012.2214050>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Yi-lun Wang received his BS degree from Yangzhou University, China, in 2015 and MS degree from Nanjing University of Science and Technology, China, in 2018. He is currently a PhD candidate in optical engineering at Nanjing University of Science and Technology. His research interests include digital low-light imaging devices and low-light-level imaging.



Yi-zheng Lang received his BS degree from China Jiliang University, China, in 2018. He is currently a PhD candidate in optoelectronics information engineering at Nanjing University of Science and Technology. His research interests include image processing, multispectral image fusion, low-light-level imaging, and their hardware implementation.



Yun-sheng Qian received his MS and PhD degrees from Nanjing University of Science and Technology, China. He is a professor of optical technology. His research interests include physical electronics and image processing.