

RESEARCH ARTICLE

MTIE-Net: Multi-technology fusion of low-light image enhancement network

Jing Tao ^{1,2}, Hao Wu^{1,2*}, Zhihao Ni^{1,2}, Zhongyang Jin^{1,2}, Changhua Zhong^{1,2}

1 Automation and Information School of Automation and Information Engineering, Sichuan University of Science & Engineering, Zigong, Sichuan Province, China, **2** Artificial Intelligence Key Laboratory of Sichuan Province, Zigong, Sichuan Province, China

* wuhao801212@163.com

Abstract

Images obtained in low-light scenes are often accompanied by problems such as low visibility, blurred details, and color distortion, enhancing them can effectively improve the visual effect and provide favorable conditions for advanced visual tasks. In this study, we propose a Multi-Technology Fusion of Low-light Image Enhancement Network (MTIE-Net) that modularizes the enhancement task. MTIE-Net consists of a residual dense decomposition network (RDD-Net) based on Retinex theory, an encoder-decoder denoising network (EDD-Net), and a parallel mixed attention-based self-calibrated illumination enhancement network (PCE-Net). The low-light image is first decomposed by RDD-Net into a lighting map and reflectance map; EDD-Net is used to process noise in the reflectance map; Finally, the lighting map is fused with the denoised reflectance map as an input to PCE-Net, using the Fourier transform for illumination enhancement and detail recovery in the frequency domain. Numerous experimental results show that MTIE-Net outperforms the comparison methods in terms of image visual quality enhancement improvement, denoising, and detail recovery. The application in nighttime face detection also fully demonstrates its promise as a pre-processing means in practical applications.

OPEN ACCESS

Citation: Tao J, Wu H, Ni Z, Jin Z, Zhong C (2024) MTIE-Net: Multi-technology fusion of low-light image enhancement network. PLoS ONE 19(2): e0297984. <https://doi.org/10.1371/journal.pone.0297984>

Editor: Sen Xiang, Wuhan University of Science and Technology, CHINA

Received: September 19, 2023

Accepted: January 14, 2024

Published: February 2, 2024

Copyright: © 2024 Tao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying the results presented in the study are available from (<https://github.com/JianghaiSCU/R2RNet>).

Funding: This research was funded by The Project of Sichuan provincial science and Technology Department (Grant No. 2022YFS0518, 2022ZHC0035); The artificial intelligence key laboratory of Sichuan province Foundation (2023RY06); Enterprise informatization and Internet of things measurement and control technology key laboratory project of Sichuan provincial university (2022WYY04); Talent

1. Introduction

With the rapid development of deep learning, digital image processing technology has been widely used in industrial production, video surveillance, daily life, military applications and other fields. However, there are frequently objective factors that lead to image defects during the image acquisition process. For instance, under weak illumination conditions, such as at night or on cloudy days, the light reflected from the surface of objects is generally weak. This often results in images that are characterized by low brightness, low contrast, color distortion, and high noise, as demonstrated in studies [1–3]. In addition, for color low-light images, the pixel values are mainly concentrated in a low range, and the gray scale difference of the corresponding pixels between each channel is also very limited, with only a small difference between the maximum and minimum gray scale of the image, a deviation of the overall color layer, and weak edge information, which leads to the difficulty in distinguishing the details of the image when observed by human beings or processed by computers. These characteristics seriously

Introduction Project of Sichuan University of Science and Engineering(2021RC12). The artificial intelligence key laboratory of Sichuan province Foundation (2019RYY01);H.W The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Talent Introduction Project of Sichuan University of Science and Engineering (2021RC12); H.W The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

reduce the subjective visual effects of low-light images, resulting in a significant reduction in their usability [4]. Enhancing low-light images and converting them into high-quality clear images can effectively improve the performance of advanced visual tasks such as target detection, multi-target tracking, semantic segmentation, underwater vision [5, 6] and face recognition. Therefore, low-light image enhancement technology has important research significance and broad application prospects as a pre-processing to improve subsequent advanced vision tasks. The purpose of low-light image enhancement is to improve its visibility, transforming the image into a form more suitable for human observation or computer processing, while suppressing noise and artifacts. At present, in order to weaken the influence of low-light images on visual tasks, researchers have proposed two solutions. One is to improve the performance of image acquisition equipment [7, 8], and the other is to perform illumination enhancement processing on the collected low-light images. Existing low-light cameras from companies such as Sony, Photonis, SiOnyx, and Texas Instruments use high-performance charge-coupled devices or complementary metal-oxide-semiconductor [9] technology, professional low-light circuits, and filters as core components to improve low-light image quality. However, due to the demanding manufacturing process, complex technology and high price of these high-performance equipment, they have not been widely used at this stage [10]. As an alternative, the use of algorithms to perform illumination enhancement on low-illumination images provides greater flexibility [11].

Over the past few decades, there has been significant progress in the research of low-light image enhancement. However, these methods are not stable in unknown application scenarios, and there is still a large room for improvement in general. For example, the noise amplification caused during the enhancement process leads to blurred details and color deviations in the enhancement results; when improving image contrast, the appropriate balance between image color, visual effects, information entropy and other factors is not considered. In summary, we propose a Multi-Technology Fusion of Low-light Image Enhancement Network (MTIE-Net). The network modularizes the enhancement task, i.e., decomposition, denoising and enhancement. Among them, the decomposition network combines dense residual blocks for global residual learning; The denoising network adopts an encoder-decoder structure, and uses channel Transformer instead of simple skip connections to transfer feature information to improve the denoising performance of the network. The enhancement network exploits the frequency information of the image. A new attentional mechanism is used for weighted fusion of different regions in an image for the purpose of enhancement. Furthermore, the enhancement process incorporates a self-calibrated training strategy. This strategy continuously adjusts the inputs at each stage, ensuring rapid convergence to an optimal level.

Overall, the contributions of this paper are as follows.

1. We propose a Multi-Technology Fusion of Low-light Image Enhancement Network (MTIE-Net). MTIE-Net contains a dense residual decomposition network based on Retinex theory (RDD-Net), a denoising network based on encoder-decoder structure (EDD-Net) and a self-calibrated enhancement network based on parallel mixed attention mechanism (PCE-Net). Among them, RDD-Net decomposes the input low-light image into lighting map and reflectance map; EDD-Net is used to process the noise in the reflectance map; finally, the lighting map is fused with the denoised reflectance map as the input of PCE-Net, and the Fourier Transform is used to perform the illumination enhancement and detail recovery in the frequency domain.
2. We study a residual dense block to construct the more efficient residual dense decomposition network by utilizing the powerful feature extraction capability of residual dense network to facilitate the feature information transfer between channels. In the design of

EDD-Net, we use subpixel convolution for upsampling to improve the resolution of the image while denoising; The channel Transformer is used instead of skip connections in the traditional encoder-decoder architecture to better capture remote dependencies and enhance contextual understanding. In addition, we propose a new attentional mechanism that enables PCE-Net to adaptively perform weighted fusion of different regions in an image to improve the enhancement effect.

3. Numerous experimental results show that MTIE-Net outperforms the comparison methods in terms of image visual quality enhancement improvement, denoising and detail recovery. The application of MTIE-Net as a pre-processing method to face detection at night further validates its application value in improving the performance of advanced visual tasks.

1.1 Related work

Existing research on low-light image enhancement categorizes methods into two primary types: traditional methods and network-based methods. Traditional methods are mainly based on histogram equalization and the Retinex theory. The use of histogram equalization to stretch the dynamic range of an image was first proposed by Pizer et al. [12] for the enhancement of low-light images. The method is computationally simple but cannot suppress noise and has limited effect on low-light image enhancement. Jobson et al. [13] were the first to apply the Retinex [14] theory to the field of low-light image enhancement, proposing to decompose the image into lighting map and reflectance map, and then perform the enhancement process. The method achieves favorable results in noise suppression and detail recovery, but due to the large computational volume of the algorithm, it is unable to process batch low-light images quickly. Later, Chen et al. [15] combined CNN with Retinex theory and proposed RetinexNet. The network contains two modules, decomposition and illumination enhancement. The decomposition module decomposes the low-light image into lighting map and reflectance map; the illumination enhancement module is responsible for enhancing the low-light image and denoising the reflectance map before outputting the reconstructed image. In recent years, deep neural networks have been widely used in the field of image enhancement with good results due to their powerful nonlinear fitting ability [16]. Jiang Hai et al. [17] proposed a novel Real-low to Real-normal Network (R2RNet) based on Retinex theory. Combining the spatial and frequency information of the image for illumination enhancement and proposing a new frequency loss function to recover more image details. Guo et al. [18] proposed Zero Reference Depth Curve Estimation (Zero-DCE), which uses low-light image enhancement as a deep network-based image-specific curve estimation task. The method takes a low-light image as input and produces a higher order curve as its output, then adjusts the dynamic range of the original image pixel by pixel based on the output curve. It achieves low-light image enhancement through intuitive and simple nonlinear curve mapping. Xu et al. [19] designed a signal-to-noise ratio (SNR) aware converter with a new self-attention module and combined it with a convolutional model through spatial variations to achieve dynamic enhancement of low-light images with a SNR prior. Wang et al. [20] proposed a flow-based low-light image enhancement method (LLFlow). By modeling the distribution of normally exposed images as an invertible network of Gaussian distributions and using this to learn a one-to-many mapping from low-light images to the distribution of normally exposed images. The method enables better regulation of light as well as suppression of noise and artifacts. Ma et al. [21] proposed a new Self-Calibrating Illumination (SCI) learning framework, which employs a cascading illumination learning process with weight sharing to deal with low-light image enhancement

tasks. Among them, the self-calibration module can greatly reduce the computational burden of the cascade mode and realize fast, flexible and stable brightening of images in real low-light scenes. Zhang et al. [22] used a dual-histogram-based iterative threshold method and a limited histogram method with Rayleigh distribution to improve the global and local contrast of the color-corrected image. After multi-scale fusion of the two, a multiscale unsharp masking strategy to further sharpen the fused image for better visual quality.

2 Methods

2.1 Network architecture

Drawing inspiration from the works of [15, 17], MTIE-Net comprises three main stages: decomposition, denoising, and enhancement. As illustrated in Fig 1: the input low-light image is firstly decomposed into lighting map and reflectance map, so as to realize the effective separation of image features and noise. Subsequently, the reflectance map is denoised, and then the denoised reflectance map is multiplied element-by-element with the lighting map, and the result obtained is used as the input to the enhancement network, which utilizes the Fourier transform to perform illumination enhancement and detail recovery in the frequency domain.

2.1.1 RDD-Net. The Retinex theory [23] proposes a basic physical law for low-light image enhancement, i.e., the visual color image I can be decomposed into reflectance map R and lighting map L , as shown in Eq (1):

$$I = R \otimes L \tag{1}$$

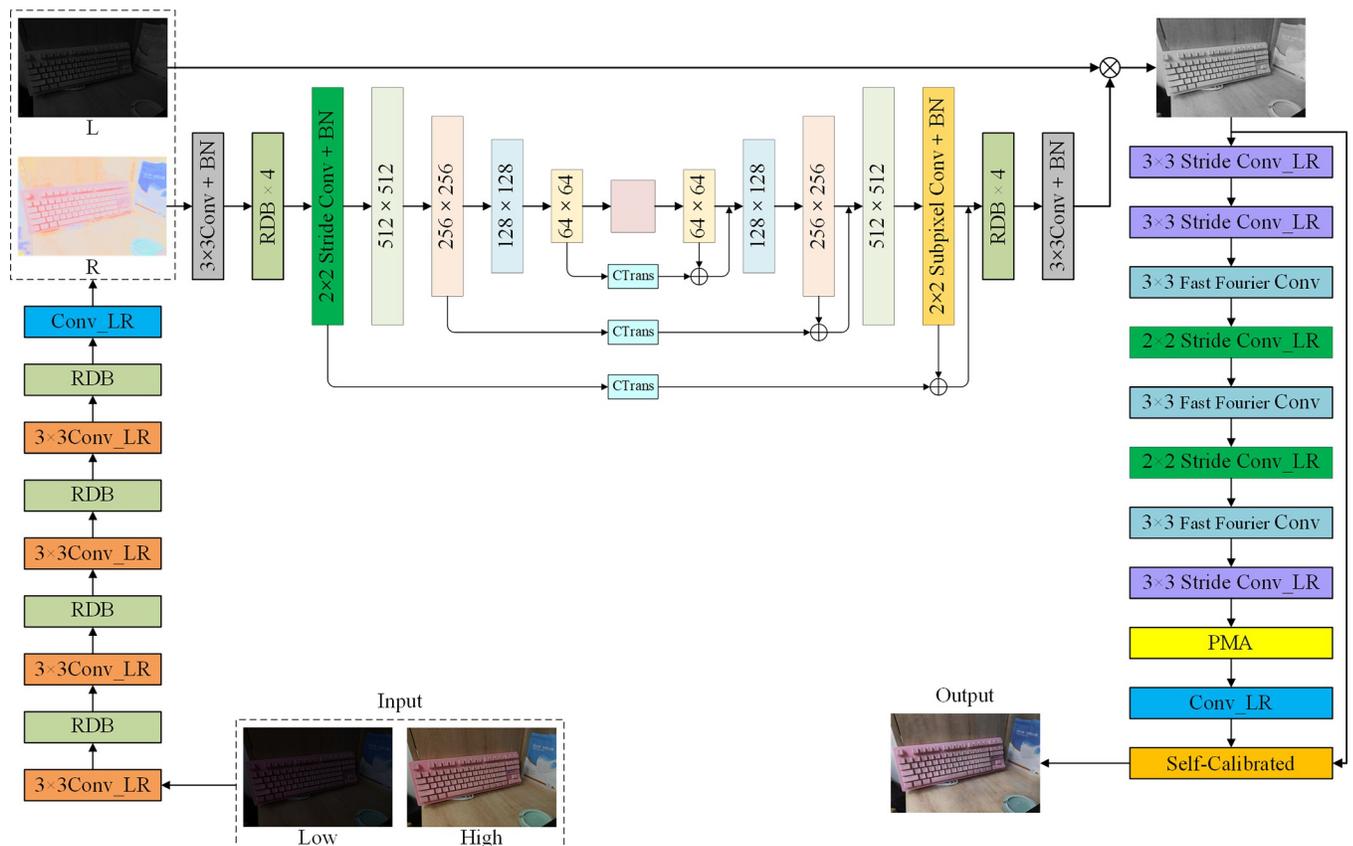


Fig 1. The network structure of the improved MTIE-Net.

<https://doi.org/10.1371/journal.pone.0297984.g001>

Where \otimes denotes element-by-element multiplication. The reflectance map is an inherent property of the image and is consistent across all luminance conditions; The lighting map represents the various luminances on an object, which are usually affected by darkness and unbalanced illumination distributions. Therefore, illumination enhancement can be achieved by changing the dynamic range of the pixels in the lighting map [15].

Retinex-based methods have been shown to be effective means of low-light image enhancement. However, in several early Retinex-based methods [13, 14, 24], the lighting map is first estimated and then the reflectance map is used as the final enhancement result. Although it can significantly restore image details, often results in unnatural and overexposed visual effects [25, 26]. In addition, existing Retinex-based decomposition networks generally control the performance through elaborate constraints and parameters, resulting in reduced model robustness across various application scenarios [15].

Studies [27, 28] have demonstrated that residual dense networks [29] and long skip connections can be used to construct deep convolutional neural networks to improve the performance of visual tasks. In order to obtain higher quality lighting maps and reflectance maps, this paper combines Residual Dense Block (RDB) with Retinex theory to construct a more efficient residual dense decomposition network (RDD-Net). Among them, the RDB consists of standard convolution with Recursive Gated Convolution (g^n Conv) [30], using LeakyReLU as the activation function. The structure of the RDB network is illustrated in Fig 2(B): firstly, shallow features are extracted using a 1×1 convolution, and the output information is adaptively controlled for local feature fusion to further improve the information flow through local residual learning; Secondly, the shallow features are used as inputs, and 3 sequential g^n Conv are used to extract the deep features, using their recursive design to improve the spatial interaction between the shallow features and the deep features; Afterwards, continuous feature transfer is generated through dense connectivity, which fully utilizes the features of all previous layers for dense feature fusion; Then a 1×1 convolution is used for global feature fusion; Finally the output feature map of the RDB is fused with the shallow features of the original input image through a dilation convolution (with a dilation factor of 2) via skip connection in order to maintain the structural similarity between the decomposed image and the original image [31]. Dilated convolution, in contrast to standard convolution, expands the sensory field without sacrificing resolution, enabling the extraction of more localized features.

2.1.2 EDD-Net. Current deep learning based low-light image enhancement methods are learned based on datasets with minimal noise [32]. However, images captured in actual low-light conditions often contain varying levels of noise, which is the main reason why such

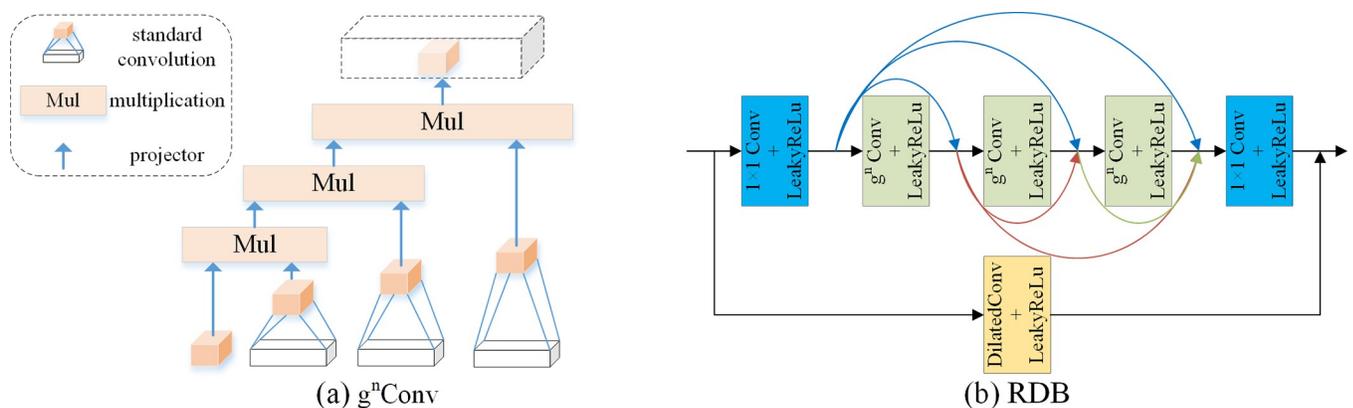


Fig 2. The network structure of RDB.

<https://doi.org/10.1371/journal.pone.0297984.g002>

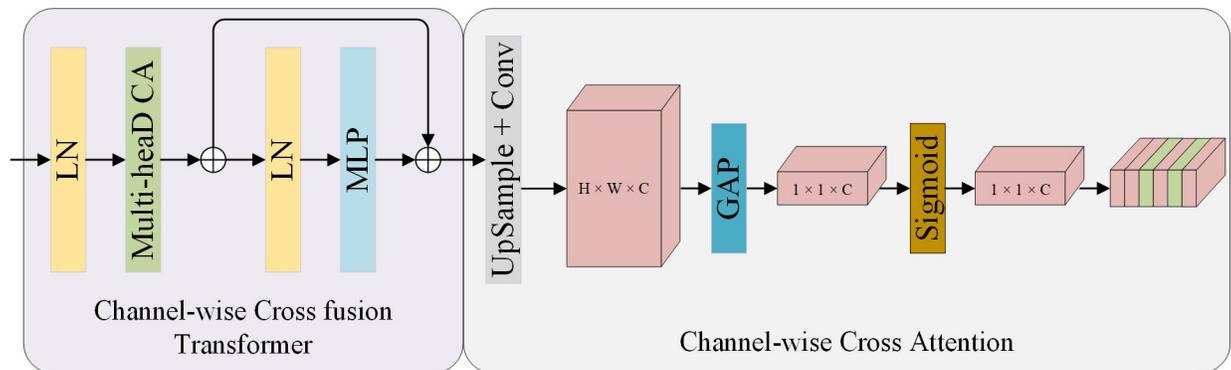


Fig 3. The network structure of CTrans.

<https://doi.org/10.1371/journal.pone.0297984.g003>

methods are usually ineffective in unknown and complex scenes. To address this, we introduce a denoising network with an encoder-decoder structure, termed EDD-Net. As a plug-and-play sub-network, EDD-Net can rapidly and flexibly handle different levels of noise generated in the enhancement process of low-light images, so that the enhancement network can be interfered by as little noise as possible in the illumination recovery process.

The backbone of EDD-Net is based on the Res_U-Net [33] architecture, which takes advantage of the high efficiency of the encoder-decoder structure for transforming images to remove noise from reflectance maps. where the encoder structure captures both low-level and high-level features; The decoder structure combines semantic features to construct the final output; The Channel Transformer (CTrans) module [34] is used to convey the spatial information lost in the pooling layer and restore the complete spatial resolution through the encoding-decoding process. As illustrated in Fig 3, CTrans consists of a Channel-wise Cross fusion Transformer and a Channel-wise Cross Attention. The former is used for multi-scale fusion of features in the encoder structure; the latter is used for fusing features that are semantically inconsistent between the decoder and decoder structures. EDD-Net operates across four scales (512, 256, 128, and 64 channels), utilizing four consecutive RDBs for up/down sampling. Redundant low-frequency information is bypassed by additional skip connections between each up/down sampling, allowing the network to focus on learning high-frequency information. In this, the downsampling process converts the noisy image from a high resolution scale to a low resolution scale using step convolution; The upsampling process uses Subpixel Convolution to transfer the low resolution scale back to the high resolution scale, thus removing noise and preserving important features of the input image.

2.1.3 PCE-Net. According to Retinex theory, the low-lighting image i is associated with the desired clear image z as follows:

$$i = z \otimes l \quad (2)$$

Where l denotes the lighting component. In general, enhancing the lighting component is regarded as the primary means of low-light image enhancement. Therefore, in this paper, we begin with the lighting map derived from RDD-Net decomposition, first fusing it with the denoised reflectance map from EDD-Net, and then in the illumination enhancement. Considering that existing low-light enhancement [18–20] focuses more on the spatial information of the image and ignores the channel information during the design, resulting in the enhanced image still having problems such as blurring of details, we study a self-calibrated illumination enhancement network based on parallel mixed attention (PCE-Net). Drawing inspiration

from [17, 35], PCE-Net initially converts the input image from the spatial to the frequency domain using the Fourier transform to obtain the frequency component information of the image. Among them, the high amplitude frequency component implies that in the spatial domain image, there is a faster change in luminance or color, which usually corresponds to information such as details, edges, textures, etc. in the image; Low amplitude frequency domain components are used to recognize the global structural information of the image, such as the overall shape, general distribution in the image, etc. Then, Fast Fourier Convolution (FFC) [36] is used to integrate the magnitude and phase information of the input image in the frequency domain and to realize the detail reconstruction by enhancing the high magnitude frequency components. Finally, the Fourier inverse transform is utilized to convert the image back to the spatial domain and output the enhanced high frequency signal.

In order to take into account the channel information of the image during the enhancement process, we study a parallel mixed attention module (PMA). As illustrated in Fig 4, PMA consists of the Pixel Attention Module (PAM) [37] fused with the Channel Attention Module (CAM) and Spatial Attention Module (SAM) mechanisms of the parallel structure, which can simultaneously attend to the channel, spatial, and pixel information of the images. In Section 3.3.1, the tandem structure of CAM and SAM suffers from distribution lag in feature extraction, which leads to an overall distortion of the color of the enhanced low-light image, whereas connecting the two in parallel focuses on both the light information and the color information of the low-light image. In addition, it was shown in [16] that low-light images have extremely high local dependencies between neighboring pixel points, While PAM can adaptively rescale the per-pixel weights for all input feature mappings. For this reason, we integrated PAM behind the parallel structure of CAM and SAM to adjust all input features channel-by-channel and pixel-by-pixel. The characterization strength of pixel information in low-light images is improved while strengthening the feature linkage between adjacent pixel points, which in turn improves the overall performance of PCE-Net.

Furthermore, PCE-Net employs a Self-Calibrated (SC) [21, 38] training strategy, which continuously calibrates the input image throughout the enhancement process to converge the results at each stage, which improves the exposure stability while greatly reducing the computational effort. Essentially, the core idea of SC is to introduce an auxiliary process in the training phase to enhance the modeling capabilities of the augmented network. This module extracts features from the input image (at the first stage) using multiple convolutions, serving

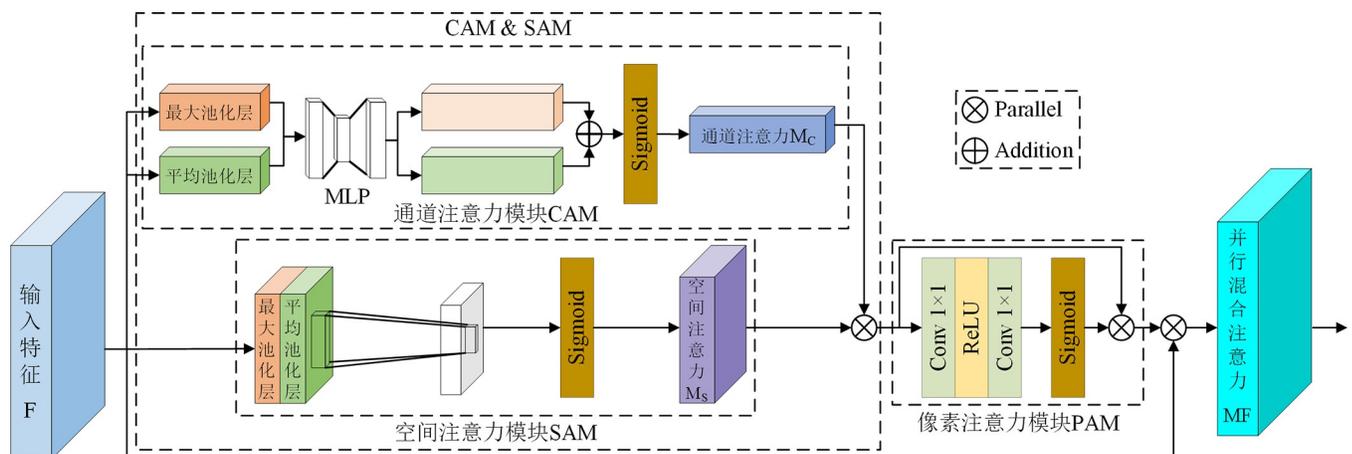


Fig 4. The network structure of PMA.

<https://doi.org/10.1371/journal.pone.0297984.g004>

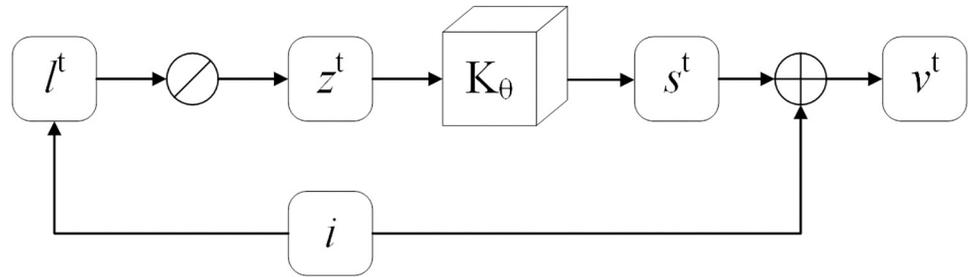


Fig 5. The calibration process.

<https://doi.org/10.1371/journal.pone.0297984.g005>

as the input for the subsequent stage of illumination estimation and generating calibration results; Connecting the inputs of each subsequent stage to the inputs of the first stage ensures that the outputs of the different stages in the training process converge to the same state, and ultimately calibrates the input images to improve the exposure stability of the model in different application scenarios. Specifically, the SC module can be represented as:

$$\mathcal{G}(I^t) : \begin{cases} z^t = I^t / I, \\ s^t = \kappa_{\theta}(z^t), \\ v^t = I + s^t, \end{cases} \tag{3}$$

where I^t is the illumination at stage t ; \emptyset denotes division by elements; κ_{θ} is the parameter learnable operator; s^t is the self-calibrated mapping at moment t ; and v^t is calibrated for the next stage of inputs. Fig 5 illustrates the calibration process in detail.

2.2 Loss function

The loss function of MTIE-Net consists of three parts [15, 17]: reconstruction loss \mathcal{L}_r , denoising loss \mathcal{L}_d , and illumination loss \mathcal{L}_i , which can be expressed as follows:

$$\mathcal{L} = \mathcal{L}_r + \mathcal{L}_d + \mathcal{L}_i \tag{4}$$

In Retinex theory, the reflectance map is computed by a pixel-by-pixel division between the input image and its lighting map mapping, which serves as a constraint on the lighting map and the reflectance map [39], the reconstruction loss \mathcal{L}_r can be expressed as follows:

$$\mathcal{L}_r = \|I - (R \otimes L)\|_1 + \|L - L_0\|_1 + \|R - I/L\|_1 \tag{5}$$

where L_0 is the initial estimated R, G, and B channel intensity maximum of the lighting map.

The primary task of EDD-Net is to recover clean signals from reflectivity maps, thus the loss function should be able to quantify the difference between the network output and the true clean signal. In this paper, smoothing L1 loss is chosen as the loss function of EDD-Net. As a balanced form of L1 loss and L2 loss, it performs better for handling both noisy data and outliers. Smoothing L1 loss adds a smoothing parameter ϵ^2 to the L2 loss, which is used to increase the stability of the values so that they do not suffer excessively as the L2 loss approaches zero. The denoising loss \mathcal{L}_d can be expressed as:

$$\mathcal{L}_d = \frac{1}{N} \sum_{j=1}^N \sqrt{((R_{nor}^j - \hat{R}_{low}^j)^2 + \epsilon^2)} \tag{6}$$

Where N is the number of samples, the index $j \in [1, N]$ represents the j th sample used to calculate the loss, which is consistent in the following equations. R_{nor}^j is the reflectance map of the reference image, and \hat{R}_{low}^j is the reflectance map of the low-light image after enhancement.

PCE-Net actually uses FFT and its inverse transform to realize the input image in the "spatial domain—frequency domain—spatial domain" between the conversion, and then achieve the purpose of illumination enhancement and detail reconstruction. Therefore, we use frequency loss [15] to enhance the performance of PCE-Net, and the frequency loss \mathcal{L}_{fre} can be expressed as:

$$\mathcal{L}_{fre} = \frac{1}{N^2} \sum_{j=1}^N inf_{f_y \sim} \prod (\hat{I}_{low}^j, I_{nor}^j) E_{(x,y)} \sim [\|\hat{I}_{low}^j - I_{nor}^j\|] \tag{7}$$

Where $inf_{f_y \sim}$ denotes joint distribution. $E_{(x,y) \sim}$ denotes using the Wasserstein distance to minimize the difference between the real part and imaginary part of enhanced result and the ground truth in frequency domain. \hat{I}_{low}^j denotes the denoised low-light image, I_{nor}^j denotes the denoised reference image.

In addition, we use the VGG loss to assist in the construction of the illumination loss, which is used to enhance the perceived quality of the enhanced image, and the VGG loss \mathcal{L}_{vgg} can be expressed as:

$$\mathcal{L}_{vgg} = \frac{1}{N} \sum_{j=1}^N (\|\phi_j(\hat{R}_{low}^j) - \phi_j(R_{nor}^j)\|^2) \tag{8}$$

Where ϕ_j denotes the loss network.

Then the illumination loss \mathcal{L}_i can be expressed as:

$$\mathcal{L}_i = \mathcal{L}_{fre} + \mathcal{L}_{vgg} \tag{9}$$

As illustrated in Fig 6, combining \mathcal{L}_r , \mathcal{L}_d and \mathcal{L}_i , MTIE-Net can converge to decompose the input image I into a reflectance map R and lighting map L ; Denoising the reflectance map and multiplying it element by element with the lighting map yields Fig 6(D); The final result is then enhanced to obtain Fig 6(E) and 6(F) shows the referenced normal light image.

3 Experiments

3.1 Implementation details

We use the publicly available dataset LSRW [17] as the training set. The performance of the model was validated using the test set of LSRW (containing 30 images), 100 randomly selected images from LOL [15], and 15 randomly selected images from the MIT-5k dataset as three independent test sets, respectively. The experimental environment configuration is shown in Table 1. For model training, the epoch is set to 40; the batch size is 4; and the patch size is 96;

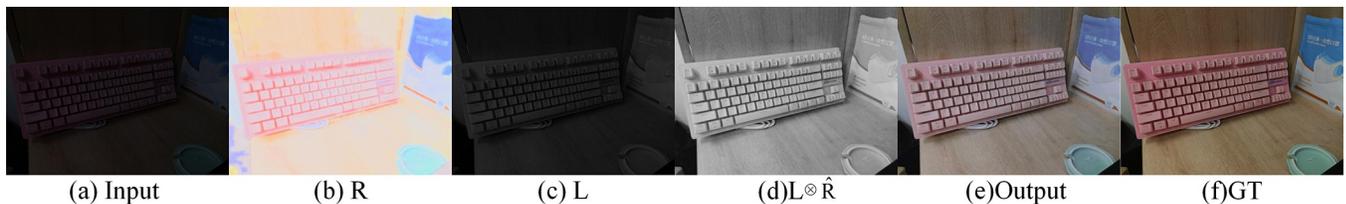


Fig 6. MTIE-Net processing and results.

<https://doi.org/10.1371/journal.pone.0297984.g006>

Table 1. Experimental setting.

Software/Hardware	Models
CPU	AMD EPYC 7302
GPU	Nvidia Ampere A100
Deep learning framework	Pytorch
Python version	3.8

<https://doi.org/10.1371/journal.pone.0297984.t001>

using the Adam [40] optimizer, the initial learning rate is set to 10^{-3} , and after 20 rounds of training the learning rate is reduced to 10^{-4} according to the learning rate decay strategy.

3.2 Evaluation metrics

In this paper, image quality assessment metrics such as peak signal-to-noise ratio (PSNR) [41], structural similarity (SSIM) [42], learned perceptual image patch similarity (LPIPS) [43], natural image quality evaluator (NIQE) [44], average gradient (AG) and spatial frequency (SF) are used to evaluate the performance of MTIE-Net. In which, PSNR, SSIM, LPIPS are full reference metrics and NIQE, AG, SF are no reference metrics. PSNR evaluates the image quality by calculating the pixel error in the global range, and the larger its value, the lower the distortion and the higher the quality of the enhanced image. SSIM combines the brightness, contrast and structural information similarity between two images to assess the quality of the processed image relative to the GT image, the size of the value between 0 and 1, the closer the value of 1 indicates that the enhanced image is more similar to the original image, the better the enhancement effect. LPIPS calculates the perceptual difference between the enhanced image and the GT image by establishing an inverse mapping relationship between the two images, with lower values indicating that the two images are more similar. NIQE predicts the perceived quality of an image by analyzing the statistical properties of the natural scene of the image, with lower values indicating that the enhanced image is of higher quality and closer to the natural image. AG is used as a measure of image sharpness, the larger the AG, the sharper the enhanced image. SF reflects the rate of change of the gray scale of the image, the larger value indicates that the image is clearer, the better the quality of the enhanced image.

3.3 Ablation study

3.3.1 PMA structural ablation study. Firstly, we conduct an ablation study on the LOL dataset to validate the effectiveness of PMA and the optimal way of integrating the three attentional mechanisms of CAM, SAM, and PAM, using PSNR and SSIM as objective evaluation metrics.

As shown in Table 2: connecting the PAM serially to the parallel structure after the CAM and the SAM, the network obtains the highest PSNR and SSIM values, achieving the best

Table 2. The ablation experiment of PMA structure.

Models	PSNR	SSIM
MTIE-Net*	17.887	0.741
MTIE-Net* + CAM + SAM	20.001	0.758
MTIE-Net* + CAM & SAM	20.033	0.765
MTIE-Net* + PAM + CAM + SAM	19.979	0.762
MTIE-Net* + CAM + SAM + PAM	19.934	0.755
MTIE-Net* + PAM + CAM & SAM	20.082	0.768
MTIE-Net* + CAM & SAM + PAM	20.250	0.771

<https://doi.org/10.1371/journal.pone.0297984.t002>



Fig 7. Comparison of subjective enhancement effect of different combined structures of CAM and SAM.

<https://doi.org/10.1371/journal.pone.0297984.g007>

enhancement results in comparison (MTIE-Net * denotes that the network has not added any of the modules of RDB, CTrans, PMA and SC; "+" denotes the serial structure and "&" denotes the parallel structure). Additionally, we randomly select two images from the LOL dataset for the enhancement test to further verify the superiority of the parallel structure of CAM and SAM. The comparison of subjective effects is illustrated in Fig 7.

3.3.2 MTIE-Net compositional ablation study. Further, to comprehensively demonstrate the positive impact of each core module in MTIE-Net, we performed ablation studies in the LOL dataset. Specifically, 1) MTIE-Net without residual dense block (-w/o RDB), 2) MTIE-Net without channel transformer (-w/o CTrans), 3) MTIE-Net without Pixel attention module (-w/o PMA), 4) MTIE-Net without self-calibrated (-w/o SC). Fig 8 can observe the visual results of the ablation experiments: (b) -w/o RDB does not reduce noise; (c) -w/o CTrans does not enhance the texture structure of images, (d) -w/o PMA does not remove color distortion, (e) -w/o SC does not improve contrast. The average scores of PSNR and SSIM for the ablation model are given in Table 3, which shows that our full model obtained the highest score. Meanwhile, it is further proved that each core module positively affects MTIE-Net.

3.3.3 Calculation cost analysis. Finally, we validate the computational efficiency of our MTIE-Net as a way to demonstrate the positive impact of the self-calibrated training strategy in reducing the computational effort and speeding up the convergence of the network. Table 4 compares the FLOPs and running time (GPU-seconds for inference) of MTIE-Net with some state-of-the-art models. To ensure the rigor of the experiments, we randomly selected 100 test images of size 600*400 from the LOL dataset and calculated their FLOPs and running time in



Fig 8. Qualitative results of ablation experiments.

<https://doi.org/10.1371/journal.pone.0297984.g008>

Table 3. Quantitative results of ablation experiments.

Model	PSRN	SSIM
MTIE-Net*	17.887	0.741
-w/o RDB	20.271	0.772
-w/o CTrans	20.109	0.769
-w/o PMA	19.874	0.766
-w/o SC	20.283	0.775
MTIE-Net (full model)	20.409	0.783

<https://doi.org/10.1371/journal.pone.0297984.t003>

Table 4. Comparison of calculation costs.

Method	FLOPs	running time
RetinexNet	136.0155	0.1198
RUAS	0.2871	0.0063
SCI	0.0724	0.0021
ZERO-DCE	5.2132	0.0051
MTIE-Net (-w/o SC)	8.2738	0.0618
MTIE-Net (full model)	2.8102	0.0045

<https://doi.org/10.1371/journal.pone.0297984.t004>

the Pytorch platform (GPU). As shown in Table 4, the FLOPs and runtime of MTIE-Net are significantly reduced after the introduction of the self-calibrated training strategy. While the computational cost of MTIE-Net is higher than SCI, notably, it is a complete network that allows for simultaneous denoising and enhancement. In conclusion, Table 4 can fully verify the efficiency and rapidity of MTIE-Net. The best result is in red whereas the second best one is in blue.

3.4 Comparison experiment

We compare MTIE-Net with RetinexNet, R2R, RUAS [45], SNR, ZERO-DCE, and SCI on three publicly available datasets.

As shown in Table 5: MTIE-Net outperforms the comparison methods in SSIM, LPIPS, NIQE, AG and SF on the LOL dataset. The average SSIM value is 0.783, the average LPIPS

Table 5. Comparison of objective evaluation metrics.

Datasets	Metrics	Method						
		RetinexNet	R2R	RUAS	SNR	ZERO-DCE	SCI	Ours
LOL	PSNR	15.720	17.892	14.846	20.476	18.059	17.253	20.409
	SSIM	0.468	0.752	0.490	0.774	0.580	0.551	0.783
	LPIPS	0.589	0.395	0.434	0.351	0.361	0.355	0.343
	NIQE	4.712	4.240	7.695	4.733	7.513	7.745	3.905
	AG	3.959	4.119	6.615	4.322	6.258	6.323	10.321
LSRW	SF	6.175	6.852	10.008	6.649	10.424	10.075	11.788
	PSNR	15.090	17.267	15.687	17.614	16.396	15.090	21.575
	SSIM	0.415	0.540	0.481	0.578	0.465	0.415	0.801
	LPIPS	0.445	0.476	0.552	0.502	0.451	0.445	0.336
	NIQE	4.035	4.408	5.800	7.546	3.820	4.035	3.921
	AG	2.517	2.664	2.504	3.933	3.843	3.826	4.675

(Continued)

Table 5. (Continued)

Datasets	Metrics	Method						
		RetinexNet	R2R	RUAS	SNR	ZERO-DCE	SCI	Ours
MIT-5k	SF	4.865	5.055	4.057	6.498	6.978	6.826	7.692
	PSNR	14.198	17.591	18.714	17.560	16.664	19.907	18.944
	SSIM	0.727	0.774	0.758	0.652	0.780	0.834	0.761
	LPIPS	0.516	0.407	0.453	0.574	0.443	0.327	0.416
	NIQE	4.529	4.261	4.587	5.133	4.101	3.963	3.904
	AG	1.673	2.981	3.598	1.117	3.299	3.184	3.247
	SF	3.657	4.824	5.297	2.649	5.932	5.289	5.810

<https://doi.org/10.1371/journal.pone.0297984.t005>

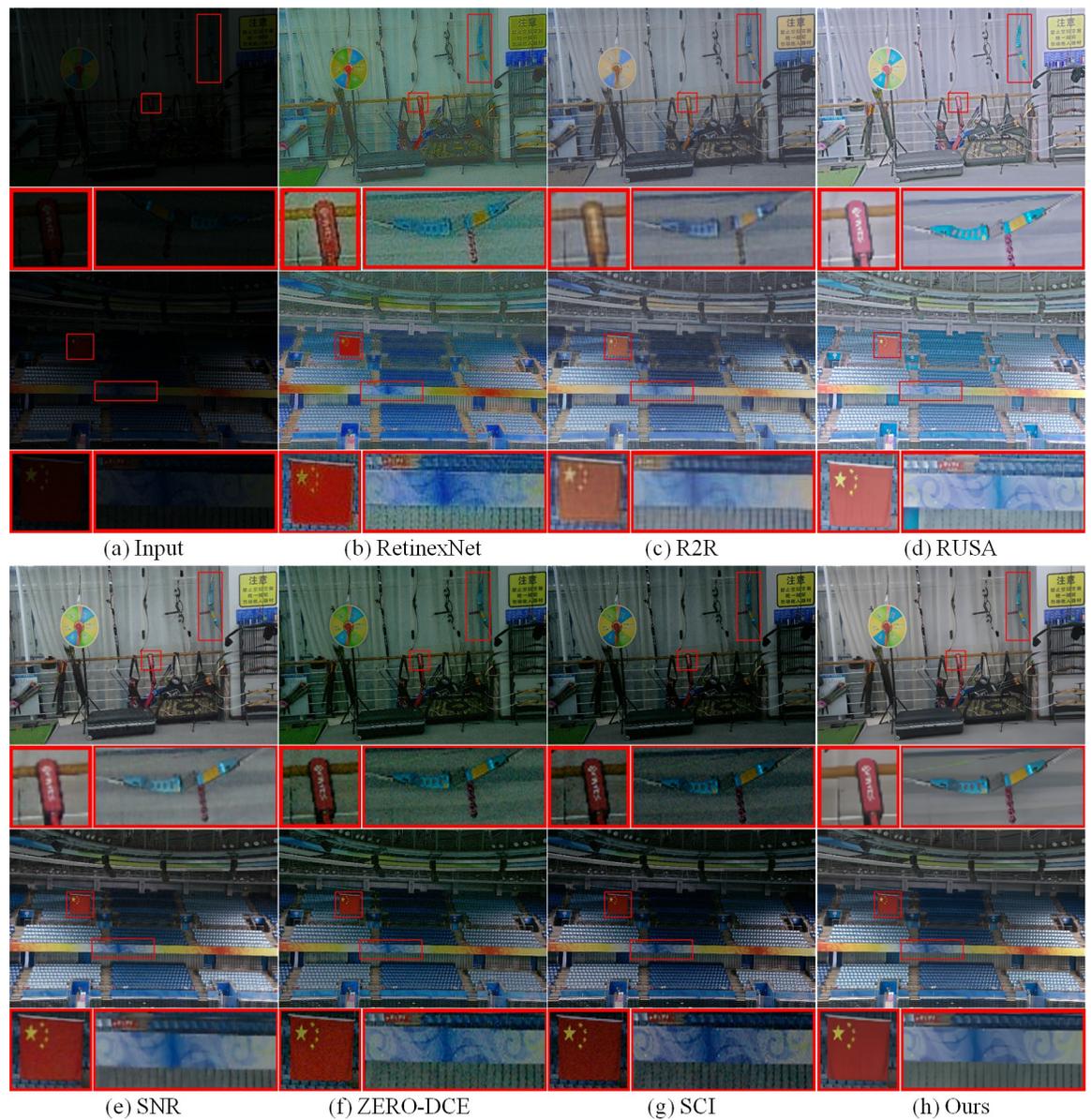


Fig 9. Comparison of enhancement effects on the LOL dataset.

<https://doi.org/10.1371/journal.pone.0297984.g009>



Fig 10. Comparison of enhancement effects on the LSRW dataset.

<https://doi.org/10.1371/journal.pone.0297984.g010>

value is 0.343, the average NIQE value is 3.905, the average AG value is 10.321 and the average SF value is 11.788; the average PSNR value is 20.409 dB, which is only 0.067 dB lower than the SNR with the highest average value. On the LSRW dataset, the PSNR, SSIM, LPIPS, AG and SF of MTIE-Net outperform the comparison methods. The average PSNR value was 21.575, the average SSIM value was 0.801, the average LPIPS value was 0.336, the average AG value was 4.675 and the average SF value was 7.692; the average NIQE value was 3.921, which was only 0.101 higher than the ZERO-DCE, with the lowest average value. On the MIT-5k dataset, MTIE-Net achieved the lowest NIQE values; the average AG value is 3.247, which is only 0.052 lower than the ZERO-DCE with the highest average value; the average SF value is 5.810, which is only 0.122 lower than the ZERO-DCE with the highest average value. In summary, it is



Fig 11. Comparison of nighttime face detection effect.

<https://doi.org/10.1371/journal.pone.0297984.g011>

shown that MTIE-Net achieves the best enhancement performance on both LOL and LSRW datasets. Comprehensive performance is also at the top of the MIT-5k dataset.

Additionally, in order to further verify the superiority of MTIE-Net enhancement performance, we randomly selected two low-light images from each of the LOL and LSRW datasets to test the enhancement effect of the above comparison algorithms.

As demonstrated in Figs 9 and 10, RetinexNet exhibits detail blurring and noise amplification on both datasets; R2R showed significant color distortion problems on both datasets; RUSA showed varying degrees of overexposure on both datasets, while ZERO-DCE and SCI showed significant underexposure, especially on the LOL dataset; Objects in SNR enhanced images have distinct colors, but the network is inferior to MTIE-Net in image detail and texture restoration. Detailed comparisons of the enhancement effects of each algorithm have been labeled with red wireframes in Figs 9 and 10. In summary, MTIE-Net shows significant enhancement performance across various low-light datasets. Compared to existing single enhancement techniques, MTIE-Net is able to enhance the luminance and preserve details of low-light images more effectively, while suppressing noise. A large number of comparison experiments in Table 4 also fully prove the effectiveness and superiority of MTIE-Net.

3.5 Pre-processing for face detection in low-light environments

In this section, we compare MTIE-Net as a pre-processing tool for nighttime face detection with the algorithm described in Section 3.4, in order to verify its impact on advanced visual tasks. We use the current state-of-the-art face detection algorithm RetinaFace [46] in the DARK FACE [47] dataset for comparison experiments. As illustrated in Fig 11: MTIE-Net improves RetinaFace detection performance most significantly in low-light environments. The experimental results further validate the generalizability of MTIE-Net and also show its potential for application in advanced vision tasks.

4 Conclusion

In this paper, we propose a Multi-technology fusion of low-light image enhancement network (MTIE-Net). The network modularizes the low-light enhancement task into three parts: decomposition, denoising and enhancement based on Retinex theory. Our experiments on several publicly available datasets demonstrate MTIE-Net's ability to effectively enhance visual quality and suppress noise in images. In addition, the nighttime face detection experiments demonstrate the promising application of MTIE-Net as a pre-processing means to enhance the performance of advanced visual tasks. Overall, this study presents a comprehensive and effective solution to low-light image enhancement, offering significant theoretical and practical value.

Supporting information

S1 File. Supporting information contains the study's minimal underlying data set. (ZIP)

Author Contributions

Conceptualization: Hao Wu, Zhongyang Jin.

Data curation: Changhua Zhong.

Formal analysis: Jing Tao.

Funding acquisition: Hao Wu.

Investigation: Jing Tao, Zhihao Ni, Zhongyang Jin.

Methodology: Jing Tao, Hao Wu, Zhongyang Jin.

Project administration: Zhihao Ni.

Resources: Jing Tao, Hao Wu.

Software: Zhihao Ni.

Supervision: Hao Wu.

Validation: Zhihao Ni.

Visualization: Zhihao Ni, Zhongyang Jin.

Writing – original draft: Jing Tao.

Writing – review & editing: Hao Wu.

References

1. Yang K F, Zhang X S, Li Y J. A biological vision inspired framework for image enhancement in poor visibility conditions[J]. *IEEE Transactions on Image Processing*, 2019, 29: 1493–1506.
2. Yang Jie, et al. "Optimization algorithm for low-light image enhancement based on Retinex theory." *IET Image Processing* 17.2 (2023): 505–517.
3. Tian Z, Qu P, Li J, et al. A Survey of Deep Learning-Based Low-Light Image Enhancement[J]. *Sensors*, 2023, 23(18): 7763.
4. Jiang Y., Gong X., Liu D., Cheng Y., Fang C., Shen X., et al. Enlightengan: Deep light enhancement without paired supervision, *IEEE Trans. Image Process.* 30 (2021) 2340–2349. <https://doi.org/10.1109/TIP.2021.3051462> PMID: 33481709
5. Zhang W, Jin S, Zhuang P, et al. Underwater image enhancement via piecewise color correction and dual prior optimized contrast enhancement[J]. *IEEE Signal Processing Letters*, 2023, 30: 229–233.
6. Zhang W, Zhou L, Zhuang P, et al. Underwater image enhancement via weighted wavelet visual perception fusion[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.
7. Jin Z H U, Li L, Wei-qi J, et al. Natural-appearance colorization and enhancement for the low-light-level night vision imaging[J]. *Acta Photonica Sinica*, 2018, 47(4).
8. Ko S, Yu S, Kang W, et al. Artifact-free low-light video enhancement using temporal similarity and guide map[J]. *IEEE transactions on industrial electronics*, 2017, 64(8): 6392–6401.
9. Yuantao Z, Mengyang C, Dexin S, et al. Digital TDI technology based on global shutter sCMOS image sensor for low-light-level imaging[J]. *Acta Optica Sinica*, 2018, 38(9).
10. Liang Z, Zhang W, Ruan R, et al. GIFM: An Image Restoration Method With Generalized Image Formation Model for Poor Visible Conditions[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1–16.
11. Chandrasekharan R, Sasikumar M. Fuzzy transform for contrast enhancement of nonuniform illumination images[J]. *IEEE signal processing letters*, 2018, 25(6): 813–817.
12. PIZER S M, MBURN E P, USTIN J D, et al. Adaptive Histogram Equalization and Its Variations [J]. *Computer Vision, Graphics, and Image Processing*, 1987, 39 (3): 355–368.
13. Jobson D J, Rahman Z, Woodell G A. Properties and performance of a center/surround retinex[J]. *IEEE transactions on image processing*, 1997, 6(3): 451–462.
14. Land E H. The retinex theory of color vision[J]. *Scientific american*, 1977, 237(6): 108–129.
15. Wei C, Wang W, Yang W, et al. Deep retinex decomposition for low-light enhancement[J]. *arXiv preprint arXiv:1808.04560*, 2018.
16. Liu G H, Yang Q, Meng Y B, et al. A progressive fusion image enhancement method with parallel mixed attention[J]. *OptoElectron Eng*, 2023, 50(4): 220231
17. Hai J, Xuan Z, Yang R, et al. R2rnet: Low-light image enhancement via real-low to real-normal network [J]. *Journal of Visual Communication and Image Representation*, 2023, 90: 103712.
18. Guo C, Li C, Guo J, et al. Zero-reference deep curve estimation for low-light image enhancement[C]// *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020: 1780–1789.
19. Xu X, Wang R, Fu C W, et al. SNR-aware low-light image enhancement[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022: 17714–17724.
20. Wang Y, Wan R, Yang W, et al. Low-light image enhancement with normalizing flow[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. 2022, 36(3): 2604–2612.
21. Ma L, Ma T, Liu R, et al. Toward fast, flexible, and robust low-light image enhancement[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022: 5637–5646.
22. Zhang W, Wang Y, Li C. Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement[J]. *IEEE Journal of Oceanic Engineering*, 2022, 47(3): 718–735.
23. Land E H, McCann J J. Lightness and retinex theory[J]. *Josa*, 1971, 61(1): 1–11.
24. Liu J, Yang W, Yang S, et al. Erase or fill? deep joint recurrent rain removal and reconstruction in videos [C]// *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018: 3233–3242.
25. Wu W, Weng J, Zhang P, et al. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement[C]// *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022: 5901–5910.

26. Li M, Liu J, Yang W, et al. Structure-revealing low-light image enhancement via robust retinex model[J]. *IEEE Transactions on Image Processing*, 2018, 27(6): 2828–2841.
27. Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017: 136–144.
28. Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks [C]//*Proceedings of the European conference on computer vision (ECCV)*. 2018: 286–301.
29. Wang X, Yu K, Wu S, et al. Esrgan: Enhanced super-resolution generative adversarial networks[C]//*Proceedings of the European conference on computer vision (ECCV) workshops*. 2018: 0–0.
30. Rao Y, Zhao W, Tang Y, et al. Hornet: Efficient high-order spatial interactions with recursive gated convolutions[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 10353–10366.
31. R D C., Kandasamy N, Rajendran S. Integration of dilated convolution with residual dense block network and multi-level feature detection network for cassava plant leaf disease identification[J]. *Concurrency and Computation: Practice and Experience*, 2022, 34(11): e6879.
32. Zhang K, Li Y, Zuo W, et al. Plug-and-play image restoration with deep denoiser prior[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(10): 6360–6376.
33. Xiao X, Lian S, Luo Z, et al. Weighted res-unet for high-quality retina vessel segmentation[C]//2018 9th international conference on information technology in medicine and education (ITME). IEEE, 2018: 327–331.
34. Wang H, Cao P, Wang J, et al. Utransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer[C]//*Proceedings of the AAAI conference on artificial intelligence*. 2022, 36(3): 2441–2449.
35. Zou W, Jiang M, Zhang Y, et al. Sdwnet: A straight dilated network with wavelet transformation for image deblurring[C]//*Proceedings of the IEEE/CVF international conference on computer vision*. 2021: 1895–1904.
36. Chi L, Jiang B, Mu Y. Fast fourier convolution[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 4479–4488.
37. Peng H, Chen X, Zhao J. Residual pixel attention network for spectral reconstruction from RGB images [C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020: 486–487.
38. Qu P, Tian Z, Zhou L, et al. SCDNet: Self-Calibrating Depth Network with Soft-Edge Reconstruction for Low-Light Image Enhancement[J]. *Sustainability*, 2023, 15(2): 1029.
39. Zhu A, Zhang L, Shen Y, et al. Zero-shot restoration of underexposed images via robust retinex decomposition[C]//2020 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2020: 1–6.
40. Kinga D, Adam J B. A method for stochastic optimization[C]//*International conference on learning representations (ICLR)*. 2015, 5: 6.
41. Sheikh H. R., Sabir M. F. and Bovik A. C., "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," in *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006, <https://doi.org/10.1109/tip.2006.881959> PMID: 17076403
42. Zhou Wang A. C. Bovik, Sheikh H. R. and Simoncelli E. P., "Image quality assessment: from error visibility to structural similarity," in *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004, <https://doi.org/10.1109/tip.2003.819861> PMID: 15376593
43. Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018: 586–595.
44. Mittal A, Soundararajan R, Bovik A C. Making a "completely blind" image quality analyzer[J]. *IEEE Signal processing letters*, 2012, 20(3): 209–212.
45. Liu R, Ma L, Zhang J, et al. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021: 10561–10570.
46. Deng J, Guo J, Zhou Y, et al. Retinaface: Single-stage dense face localisation in the wild[J]. *arXiv preprint arXiv:1905.00641*, 2019.
47. Yang W, Yuan Y, Ren W, et al. Advancing image understanding in poor visibility environments: A collective benchmark study[J]. *IEEE Transactions on Image Processing*, 2020, 29: 5737–5752.