

Received 13 August 2022, accepted 24 August 2022, date of publication 29 August 2022, date of current version 9 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3202940

RESEARCH ARTICLE

Low-Light Image Enhancement via Gradient Prior-Aided Network

YUXU LU^{1,*}, YUAN GAO^{1,*}, YONGQI GUO², WENYU XU³, AND XIANJUN HU⁴

¹School of Navigation, Wuhan University of Technology, Wuhan 430063, China

²Personnel Office, Wuhan University of Technology, Wuhan 430070, China

³Wuhan Baosight Software Company Ltd., Wuhan 430080, China

⁴College of Electronic Engineering, Naval University of Engineering, Wuhan 430033, China

Corresponding authors: Yongqi Guo (yqguo@whut.edu.cn) and Wenyu Xu (xuwenyu_971652@baosight.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 72004174 and Grant 51979215.

*Yuxu Lu and Yuan Gao are co-first authors.

ABSTRACT Low-light images have low brightness and low contrast, which brings huge obstacles to the intelligent video surveillance system. The enhancement of low-light images must simultaneously consider the interference of factors such as brightness, contrast, artifacts, and noise. To this end, in this study, we propose a gradient prior-aided low-light enhancement network (GPANet). The main idea is to improve the network's ability to extract edge features and remove unwanted noise by introducing first-order (i.e., Sobel Filter) and second-order gradient (i.e., Laplacian Filter) features. Unlike in previous methods, in the proposed study, we first extract the first-order and second-order gradient information of low-light images and concatenate them with low-light images for multi-view feature analysis in the multi-view fusion encoder (MFE). Then, we suggest the multi-branch topology module (MTM) to fuse and decompose the multi-view features. Finally, we reconstruct the multi-view features through multi-view decomposition decoders (MDDs, including three sub-decoders) to generate potentially normal-light images. The first- and second-order gradient decoders will provide the enhancement decoder with multi-scale gradient prior features. Furthermore, we suggest a residual network to speed up network convergence while ensuring stable enhancement performance. We conduct experiments on widely adopted datasets. The results demonstrate the advantages of our method compared to other methods from both qualitative and quantitative perspectives. The source code is available at <https://github.com/LouisYuxuLu/GPANet>.


INDEX TERMS Gradient prior-aided, low-light image enhancement, multi-branch topology, multi-view features, residual network.

I. INTRODUCTION

Computer vision provides sensing and processing technologies for various applications and services, including the current vision-based unmanned driving system and visible light-based intelligent surveillance system. However, insufficient light in low-light conditions has an effect on the initially visible scene information. The light reflection of the scene perceived by the photosensitive device on the camera is weak, the captured image is dim and blurred, and the image has unwanted noise interference, etc. This visual deterioration severely restricts the development of advanced visual tasks (such as object detection [1], [2] and segmentation [3]) and

causes an unpleasant visual experience. Therefore, low-light image enhancement is essential for improving the performance of existing optical systems.

In recent decades, to extract the hidden feature information from the low-light background, researchers have proposed a variety of solutions for low-light image enhancement. Methods for low-light image enhancement can be divided into two categories: traditional methods and learning-based methods. Histogram equalization (HE)-based, Retinex-based, and Dehazing-based are the main solutions for traditional enhancement methods. HE-based methods [4], [5], [6], [7] adjust the contrast and brightness of low-light images by increasing the gray level and flattening the intensity or color distribution of pixels. Due to the difficulty of adjusting the gray level adaptively, the enhanced image will

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan .

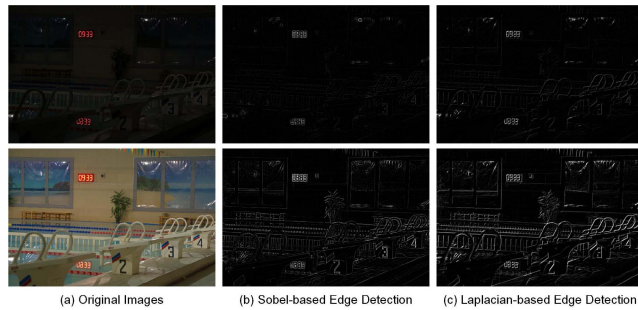


FIGURE 1. Sobel- [37] and Laplacian-based [38] edge detection results for low-light and normal-light images from LOL dataset [25].

experience the phenomenon of the loss of local details or an exaggerated increase in contrast. Retinex-based methods [8], [9], [10], [11], [12], [13], [14], [15], [16] decompose the low-light image into illumination and reflection components. The illumination component is then analyzed and adjusted to generate a potential normal-light image. However, Retinex-based methods are susceptible to local color distortion and cannot efficiently eliminate noise. Dehazing-based methods [17], [18], [19] have also been extensively attempted for low-light image enhancement. The reversed low-light image can be dehazed [20] and then reversed to generate an enhanced image. However, the pixel value composition of the inverted low-light and hazy images are significantly different, so Dehazing-based methods are susceptible to localized color distortion. The camera response model-based methods [17], [21], [22] adjust the brightness of the image by adjusting each pixel to the desired exposure based on an estimated exposure ratio map. Deep learning [23] has had a great deal of success with low-level computer vision tasks. Learning-based methods [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36] have been extensively studied in low-light image enhancement. It mainly comprises end-to-end learning, model-based learning, and generative adversarial learning. However, the learning-based methods are challenging to meet the enhancement requirements of different low-light scenes. In addition, it will inevitably cause the loss or destruction of detailed information, such as the edge texture of the image.

Real-world low-light images will mask or destroy potential edge texture features due to insufficient brightness, noise interference, *et al.*, resulting in local over-smoothing or serious noise interference in the enhanced image. Considering the advantages and disadvantages of the above methods, to more accurately extract the latent edge feature information in the low-light background, we propose to introduce first- (i.e., Sobel Filter [37]) and second-order (i.e., Laplacian Filter [38]) gradient features [39], [40] to assist the reinforcement network in learning the mapping of global and local edge features. The first- and second-order spatial derivatives are beneficial to improve the stability of the network for edge information extraction and the robustness of noise suppression. To this end, we propose a gradient prior-aided low-light enhancement network (GPANet). Specifically, the

multi-view fusion encoder (MFE), multi-branch topology module (MTM), and multi-view decomposition decoders (MDDs) enable the transmission and exchange of three-view feature information (i.e., Sobel-, Laplacian-based, and original) at three scales. In addition, through multi-view and multi-scale feature information fusion and decomposition, our GPANet can capture the imperceptible edge detail features from the dark background more accurately, thereby producing a potential normal-light image. The main contributions of our work can be summarized as follows

- We introduce first-order Sobel and second-order Laplacian gradient priors to assist the learning and mapping of edge features in low-light images by the deep enhancement network.
- We construct an end-to-end network consisting of MFE, MTM, MDDs. It can improve the enhancement performance of the deep network by fusing and decomposing multi-view and multi-scale features.
- Extensive quantitative and qualitative evaluation experimental results demonstrate that GPANet can achieve high-quality low-light image enhancement in complex imaging environments compared to state-of-the-art methods.

The rest of this paper is organized as follows. The recent studies on low-light image enhancement are reviewed in Section II. In Section III, we introduce our GPANet. Numerous experiments on both synthetic and real-world scenarios have been implemented to evaluate the enhancement performance in Section IV. Conclusions and discussion are given in Section V.

II. RELATED WORK

In this section, we briefly review the research on low-light image enhancement methods, including traditional and learning-based methods.

A. TRADITIONAL LOW-LIGHT ENHANCEMENT METHODS

1) HE-BASED

The histogram equalization (HE) [4] aims to control the processed image histogram, enabling the pixel values to follow a uniform distribution, which improves the contrast and clarity of the image. Because it is a global operation that ignores the brightness transformation, it will result in over or under enhancement issues. The dynamic histogram equalization (DHE) [41] separates the histogram into sub-histograms and equalizes each one. Contrast limit adaptive histogram equalization (CLAHE) [42] adaptively controls the degree of HE contrast enhancement. The method described above may cause major color misalignment issues, and the details in darker areas are often not appropriately enhanced. To enhance the overall visual effect, some succeeding methods have improved the HE by maintaining the average value of image brightness and enhancing resilience against noise, etc. For example, brightness preserving dynamic histogram equalization (BPDHE) [43] is

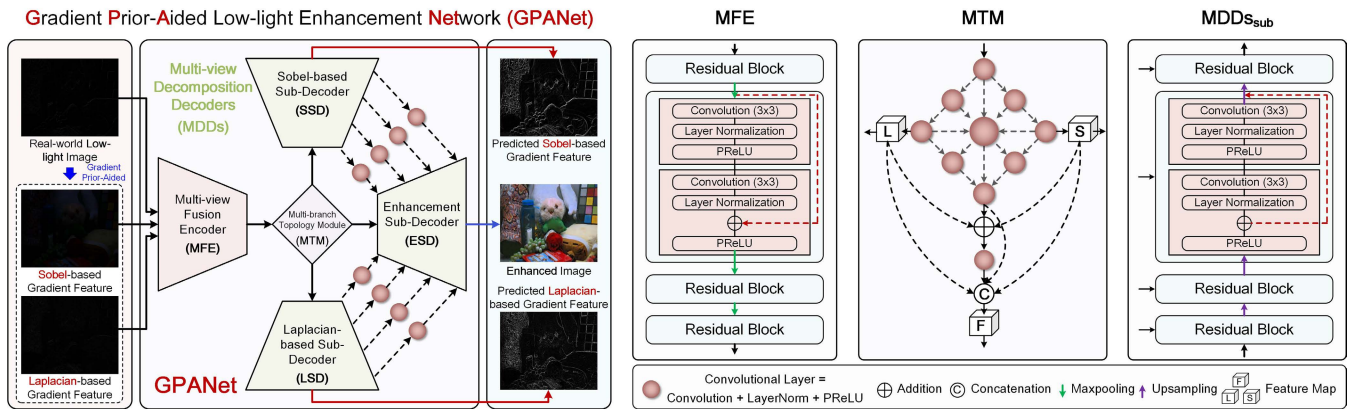


FIGURE 2. The flowchart of our gradient prior-aided low-light enhancement network (GPANet). It consists of three modules, including Multi-view fusion encoder (MFE), Multi-branch topology module (MTM), and Multi-view decomposition decoders (MDDs).

appropriate for images with a very dark foreground and background because it improves image contrast while maintaining the original image’s overall brightness. Nonetheless, it is probably to increase background noise, limit proper signal contrast, and generate excessive saturation in some parts of the image. Lee *et al.* [44] also improved contrast by increasing the gray value of adjacent pixels. Shubhi *et al.* [45] proposed a novel unsharp mask filtering technique combined with histogram equalization to maximize the entropy of generic images. In addition, it controls over-enhancement and under-enhancement by cropping the histogram of the image. Nonlinear exposure intensity-based modification histogram equalization (NEIMHE) [46] divides a non-uniformly illuminated image into five sub-regions and modifies each sub-region histogram by setting nonlinear weights in the cumulative density function of each sub-region histogram. HE-based methods effectively enhance contrast in the whole or part of the image, but most methods are inflexible. Some parts of the image have undesirable visual effects, such as underexposure, overexposure, and amplified noise.

2) RETINEX-BASED

The principle of the Retinex theory [8] is to divide the image into the illumination component and the reflection component and then restore the original detail information of the image. After estimating and boosting the illumination component, the two images are fused to achieve an enhanced effect. The single-scale Retinex method (SSR) [47] approximated the reflection component by the Gaussian function and the image convolution. The multi-scale Retinex restoration (MSR) [48] was a modification of SSR. It used Gaussian filtering at various scales to approximate the illumination image, followed by a weighted average of the filtering results. The multi-scale Retinex with color restoration (MSRCR) [49] was built upon MSR and included a color restoration factor to correct image distortion and bring it closer to the real scene. After that, several methods emerged, which combined the Retinex theory and other theories. For example,

structure-revealing low-light image enhancement (SRIE) [11] employed a compensation technique to compensate for the dark areas of the image that are exaggerated by the logarithmic domain gradient. Low-light image enhancement (LIME) [12] estimated the brightness component and used an inverse technique to obtain the reflection component. Naturalness preserved enhancement (NPE) [10] employs Retinex theory and log bilateral conversion to bring the light component mapping closer to natural color. To better remove unwanted noise in low-light images, the low-rank regularized retinex model (LR3M) [50] injected a low-rank prior to the retinex decomposition process. It can avoid remaining noise commonly found in illumination and reflectance maps by sequentially estimating piecewise smooth illumination and noise-suppressed reflectance. Wang *et al.* [51] proposed an improved logarithmic transformation-based adaptive and simple color image enhancement method by applying the Weber-Fechner law to grayscale mapping in logarithmic space. The above Retinex-based methods improve image contrast and mitigate the effect of noise to a certain extent.

3) DEHAZED-BASED

Due to the similar properties of low-light and hazy images, mature dehazing methods [20] can be used to enhance low-light images. Dong *et al.* [17] enhanced the image by first inverting the low-light image and then improving the image’s contrast by the dehazing method. Li *et al.* [19] employed an appropriate BM3D [52] denoising procedure to separate the base and enhancement layers and then adjusted the two layers independently. While these methods provide respectable findings, they lack a physically plausible explanation.

B. LEARNING-BASED LOW-LIGHT ENHANCEMENT METHODS

Convolutional neural networks (CNN) have been widely employed for image enhancement in low light conditions. LLNet [24] is a deep learning method and trains a stacked sparse noise reduction autoencoder based on synthetic data.

Algorithm 1 Gradient Prior-Aided Network (GPANet)

Require: Low-light image x and its Sobel-based gradient feature G_x^s and Laplacian-based gradient feature G_x^l , backbone Network $\mathcal{F}_{\text{GPANet}}$, Learning loss $\mathcal{L}_{\text{GPANet}}$, groundtruth of three-view y , G_y^s , and G_y^l , batch size N , epoch $E = 60$.

```

1:  $\mathcal{F}_{\text{GPANet}} \rightarrow \{\mathcal{F}_{\text{MFE}}, \mathcal{F}_{\text{MTM}}, \mathcal{F}_{\text{MDDs}}\};$ 
2:  $\mathcal{F}_{\text{MDDs}} \rightarrow \{\mathcal{F}_{\text{SD}}, \mathcal{F}_{\text{LD}}, \mathcal{F}_{\text{ED}}\};$ 
3:  $\mathcal{L}_{\text{GPANet}} \rightarrow \{\mathcal{L}_{\text{Sobel}}, \mathcal{L}_{\text{Lap}}, \mathcal{L}_{\text{En}}\};$ 
4: while  $i < E$  do
5:   for sampled minibatch  $\{x_k\}_{k=1}^N$  do
6:     Data pre-processing for task  $\mathcal{F}_{\text{GPANet}}$ 
7:      $y_e = \mathcal{F}_{\text{MFE}}(\text{cat}[x, G_x^s, G_x^l])$ 
8:      $y_m^s, y_m^e, y_m^l = \mathcal{F}_{\text{MTM}}(y_e)$ 
9:      $\hat{G}_y^s = \mathcal{F}_{\text{SD}}(y_m^s)$ 
10:     $\hat{y} = \mathcal{F}_{\text{ED}}(y_m^e)$ 
11:     $\hat{G}_y^l = \mathcal{F}_{\text{LD}}(y_m^l)$ 
12:    Computer gradient with respect to  $\mathcal{L}_{\text{GPANet}}$ 
13:     $\mathcal{L}_{\text{Sobel}} = \mathcal{L}_{\text{Sobel}}(\hat{G}_y^s, G_y^s)$ 
14:     $\mathcal{L}_{\text{Lap}} = \mathcal{L}_{\text{Lap}}(\hat{G}_y^l, G_y^l)$ 
15:     $\mathcal{L}_{\text{En}} = \mathcal{L}_{\text{En}}(\hat{y}, y)$ 
16:    Update layers within  $\mathcal{F}_{\text{GPANet}}$ ,
17:   end for
18: end while
19: return  $\mathcal{F}_{\text{GPANet}}$  and  $\hat{y}$ , and discard  $\hat{G}_y^s$  and  $\hat{G}_y^l$ .

```

The method can enhance low-light images while simultaneously reducing noise. HDRNet [53] learns local, global, and content-dependent decisions to approximate the intended image transformation. RetinexNet [25] builds upon the Retinex theory and adjusts the brightness map to improve low-light images using a product neural network. LightenNet [54] takes a weakly illuminated image as input and outputs its illumination map, which is then used to generate the enhanced image using the Retinex model. DeepUPE [55] improves low-light images by predicting the brightness map but does not take noise into consideration. While learning methods outperform traditional methods, most of them are ineffective at noise reduction. Moreover, some even neglect noise reduction entirely. MBLLN [26] applies multiple levels of feature extraction and fusion to low-illumination image enhancement, achieving a more noticeable enhancement effect. KinD [27], and KinD++ [56] use the Retinex theory to optimize the decomposition, while the reconstruction structure incorporates an adjustment network and efficiently performs continuous light map modification. Jiang et al. [29] proposed EnlightenGAN that employs a global-local discriminator and a self-regular attention mechanism to avoid training on paired image data sets, thereby increasing the adaptability of the network to most real-world scenarios. Guo et al. [30] established a non-reference network and developed a new zero-reference depth curve estimation method (Zero-DCE). The non-reference network addresses the over-fitting problem and has superior generalization

power to the reference network. However, it could be improved in terms of handling noise and color aberrations. Zhao et al. [57] proposed a novel Retinex decomposition “generative” strategy to generate more accurate latent components and used a unified depth framework to perform low-light image enhancement. Zhang et al. [32] proposed a novel method to learn and infer motion field (optical flow) from a single image and synthesize short-range video sequences, thereby enforcing the temporal stability in low-light video enhancement with only static images. Sobashi et al. [35] proposed a low light homomorphic filtering network, which performs image-to-frequency filter learning and is jointly trained to optimize image enhancement and classification performance. Light channel enhancement network (LiCENet) [58] suggested a combination of an autoencoder and a convolutional neural network in the HSV color space to train a low-light enhancer and further improve the details of low-light images on top of improved lighting. However, the learning-based methods have insufficient ability to express image features at different scales, making it difficult for the network to recover detailed information from extremely dark images. In addition, the enhanced image is prone to color distortion, amplified noise, and blurred edges.

III. GPANet: GRADIENT PRIOR-AIDED NETWORK

In this section, we elaborate on the details of GPANet. First, we extract Sobel- and Laplacian-based edge gradient features from low-light images. Then, we performed a multi-view fusion encoder (MFE), multi-branch topology module (MTM), and multi-view decomposition decoders (MDDs) on low-light images. Finally, we elaborate on the network parameter settings of the loss function.

A. EDGE GRADIENT DETECTION

In image processing, an edge is an intentionally abrupt change in intensity. However, low-light images have inconspicuous edge features due to inconsistent local brightness in Fig. 1. Therefore, we will learn to optimize the gradient features of low-light images to extract the masked gradient edge information in complex low-light imaging environments. Spatial edge detection techniques are mostly based on the first-order derivatives (e.g., Sobel) and the second-order derivatives (e.g., Laplacian).

1) SOBEL FILTER

The Sobel operator is a discrete difference operator that is used to approximate the image brightness function gradient. It can return the gradient vector or normal vector corresponding to any point in the image. In addition, the Sobel operator can assist in balancing edge detection and noise suppression. Considering a digital image $I(u, v)$, where (u, v) represent spatial coordinates, the image gradient magnitude is defined as

$$|G_s| = \sqrt{G_{su}^2 + G_{sv}^2} \simeq |G_{su}| + |G_{sv}|, \quad (1)$$

where G_{su} and G_{sv} are the gradient components respectively in u and v directions

$$G_{su} = \frac{\partial I(u, v)}{\partial u}, \quad \text{and} \quad G_{sv} = \frac{\partial I(u, v)}{\partial v}. \quad (2)$$

G_{su} and G_{sv} are obtained by filtering the image with directional kernels, K_{su} and K_{sv} , i.e.,

$$K_{su} = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix}, \quad K_{sv} = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \quad (3)$$

Therefore, $G_{su} = I(u, v) * K_{su}$ and $G_{sv} = I(u, v) * K_{sv}$, where the $*$ symbol denotes convolution. However, because the Sobel operator does not process the image based on its grayscale, it is challenging to distinguish the image's main body from its background, and the extracted image contour is often unsatisfactory.

2) LAPLACIAN FILTER

The Laplacian operator is a second-order derivative operator that will produce a steep zero-crossing at the edge. The Laplacian operator is isotropic and can sharpen boundaries and lines in any direction, with no directional characteristics. In mathematical terms, the Laplacian filter represents the differential operator. The continuous Laplacian of a two-dimensional function $I(u, v)$ is defined as

$$G_l = \frac{\partial^2 I(u, v)}{\partial u^2} + \frac{\partial^2 I(u, v)}{\partial v^2}. \quad (4)$$

The Laplacian filter can be represented as a 3×3 mask, with the center value being negative or positive, depending on the neighboring values. The kernel of the Laplacian filter can be given by

$$K_l = \begin{bmatrix} 0 & +1 & 0 \\ +1 & -4 & +1 \\ 0 & +1 & 0 \end{bmatrix}. \quad (5)$$

We calculate the convolution between K_l and the image $I(u, v)$ to obtain the second order gradient G_l , i.e., $G_l = I(u, v) * K_l$. The Laplacian operator method is sensitive to noise. However, considering that this paper uses Laplacian-based features as a priori information and uses deep networks for learning and prediction, it will not cause additional negative effects.

B. MULTI-VIEW FUSION ENCODER

As shown in Fig. 2, we take the low-light image x , the corresponding concatenation of gradient features G_s , and G_l together as the input (i.e., $E_{in} = \text{cat}[x, G_x^s, G_x^l]$) to the encoder, where $\text{cat}[\cdot]$ represents the concatenation of multi-view feature map. On the one hand, the multi-view feature can make the encoder purposefully pay attention to the edge gradient information in low-light images. On the other hand, the encoder can improve its image feature mining ability in complex imaging environments by receiving different types of inputs, thereby further optimizing the deep model.

Encoder concatenates several residual units (ResUnits). The particular operation in ResUnit, denoted by $R(\cdot)$, can be expressed as

$$f_{i+1}^k = R(f_i^k) = \tau \left(l \left(c \left(\tau \left(l \left(c \left(f_i^k \right) \right) \right) \right) \right) + f_i^k \right), \quad (6)$$

where f_i^k and f_{i+1}^k are the input and output of the $(i+1)$ -th ResUnit in the k -th residual module. $c(\cdot)$ is convolutional layer, $l(\cdot)$ is Layer Normalization, and $\tau(\cdot)$ is PReLU. There are 4 ResUnits in the encoder, and they are connected through the maximum pooling $M(\cdot)$. The output f_E after a sequence of operations of ResUnits is

$$f_E = R_4^E \left(M \left(R_3^E \left(M \left(R_2^E \left(M \left(R_1^E \left(E_{in} \right) \right) \right) \right) \right) \right) \right). \quad (7)$$

C. MULTI-BRANCH TOPOLOGY MODULE

The output of the encoder contains the low-light image, Sobel- and Laplacian-based large unordered, but low-dimensional feature maps that contain a lot of information. To enhance the three-view feature information, as shown in Fig. 2, we propose a multi-branch topology module (MTM). MTM can transmit information in the depth and width directions of the network through different network nodes so as to more effectively utilize the parameters between neurons and improve the network learning ability [59]. The elements that make up the MTM (i.e., the pink circles) are convolutional units, which are still composed of a convolution, a layer normalization, and the PReLU function. It is worth noting that although multi-view feature addition can enhance high-frequency information such as gradient edges, at the same time, unwanted noise is also enhanced. To this end, we will adopt a strategy of first addition fusion and then concatenation fusion of the output feature maps of the three fields of view from the MTM. In addition, we propose inserting a convolutional layer between addition and concatenation operations to fuse and enhance edge texture features.

MTM will output feature maps of three views, including Sobel-based feature y_m^s , Laplacian-based feature y_m^l , and mainline enhancement feature y_m^e . To better fuse and decompose the features of the three perspectives, we use the form of dense connections to strengthen the transfer of information flow. In addition, to improve the fusion processing capability of the convolutional layer in the middle of the MTM for huge amounts of information, we additionally increase the number of convolutional channels.

In the end, we fuse the outputs of the three views of the MTM again to strengthen the mainline features, which can be given by

$$f_M = \text{cat} \left[\tau \left(l \left(c \left(y_m^e + y_m^s + y_m^l \right) \right) \right), y_m^e, y_m^s, y_m^l \right], \quad (8)$$

where f_M is the input of Main-Decoder.

D. MULTI-VIEW DECOMPOSITION DECODER

Our GPANet will supervise the optimization of deep networks from three views. Therefore, the proposed network consists of three sub-decoders (i.e., Sobel-,Laplacian-based,

and enhancement decoders). Similar to the encoder, a decoder can be given as

$$f_D = R_4^D \left(U \left(R_3^D \left(U \left(R_2^D \left(U \left(R_1^D \left(D_{in} \right) \right) \right) \right) \right) \right) \right), \quad (9)$$

where U is upsampling operation. In this paper, we suggest using bilinear interpolation to improve the resolution of feature maps. D_{in} are the three-view outputs of the MTM, i.e., y_m^s , y_m^l , and y_m^e . In addition, we introduce long skip connections to connect residual units of corresponding scales of the encoder and decoder for more efficient learning and inference. For the enhancement decoder, each of its residual units will simultaneously receive the output of the corresponding residual units of Sobel- and Laplacian-based decoders, thereby further improving the mainline encoder’s attention to edge feature information.

E. LOSS FUNCTION

To optimize trainable parameters and improve image quality qualitatively and quantitatively in the proposed GPANet, we propose a loss function \mathcal{L}_{total} , which is composed of three components, i.e., Sobel-based gradient consistency loss \mathcal{L}_{Sobel} , Laplacian-based gradient consistency loss \mathcal{L}_{Lap} , and Final Enhancement loss \mathcal{L}_{En} . It is expressed as

$$\mathcal{L}_{total} = \omega_1 \mathcal{L}_{En} + \omega_2 \mathcal{L}_{Sobel} + \omega_3 \mathcal{L}_{Lap}, \quad (10)$$

where ω_1 , ω_2 , and ω_3 denote the weight value for each loss term, respectively. Based on extensive experiments, those values are set to 0.8, 0.1, and 0.1 for the best performance. The details of the three loss functions are given below.

1) SOBEL-BASED GRADIENT CONSISTENCY LOSS

To better approximate the image gradients, we compute the first and second derivatives and use them as regularizers (i.e., ℓ_1 and ℓ_2) to penalize incorrect estimates of image gradients, i.e., \mathcal{L}_S and \mathcal{L}_L . In addition, ℓ_1 and ℓ_2 can also improve the edge detection accuracy of the two gradient prior-Aided branch networks by forcing the semantic properties in the low-light image, making it as close to the ground truth as possible, which can be given by

$$\mathcal{L}_{Sobel} = 0.5 \cdot \left\| \hat{G}_y^s - G_y^s \right\|_1 + 0.5 \cdot \left\| \hat{G}_y^s - G_y^s \right\|_2. \quad (11)$$

2) LAPLACIAN-BASED GRADIENT CONSISTENCY LOSS

Similar to \mathcal{L}_{Sobel} , we still suggest ℓ_1 and ℓ_2 as the loss function of \mathcal{L}_{Lap} , i.e.,

$$\mathcal{L}_{Lap} = 0.5 \cdot \left\| \hat{G}_y^l - G_y^l \right\|_1 + 0.5 \cdot \left\| \hat{G}_y^l - G_y^l \right\|_2. \quad (12)$$

3) ENHANCEMENT LOSS

Whether the structure, brightness, color, etc. of the enhanced image are natural is an important criterion for testing the performance of the enhancer. To optimize the trainable parameters of the proposed method, we formulate a multi-constrained loss function \mathcal{L}_{total} , which consists of two parts, i.e., data loss \mathcal{L}_d , and edge loss \mathcal{L}_e . It is defined as

$$\mathcal{L}_{En} = \omega_1^1 \mathcal{L}_{data} + \omega_1^2 \mathcal{L}_{edge}, \quad (13)$$

in this part, we set $\omega_1^1 = 0.99$, and $\omega_1^2 = 0.01$, respectively.

TABLE 1. The details of train/test datasets used in our experiments.

Datasets	Total	Training	Testing	Ground Truth
DICM [44]	69	—	69	✗
ExDARK [60]	7363	—	60	✗
GLADNet [61]	5000	1500	0	✓
LIME [12]	10	—	10	✗
LOL [25]	1500	1485	15	✓
MEF [62]	17	—	17	✗
MIT-Adobe FiveK [63]	5000	—	250	✗
NPE [10]	85	—	85	✗
TMDIED	222	—	222	✗

From Eq. (10) and (11), we can know that ℓ_1 and ℓ_2 are able to penalize the enhancement results for not being similar to the corresponding ground truth in the pixel-averaged point of view, so we thus suggest ℓ_1 and ℓ_2 as our major data constraints loss, defined as follows

$$\mathcal{L}_{data} = 0.5 \cdot \left\| \hat{y} - y \right\|_1 + 0.5 \cdot \left\| \hat{y} - y \right\|_2. \quad (14)$$

Sobel- and Laplacian-based decoders provide the enhancement decoder with a large number of edge features of different scales. To further improve the learning and reasoning ability of the enhancement decoder for high-frequency edge features, we suggest using edge loss to constrain the difference between y and \hat{y} , i.e.,

$$\mathcal{L}_{edge} = \sqrt{(Lap(\hat{y}) - Lap(y))^2 + \varepsilon^2}, \quad (15)$$

where $Lap(\hat{y})$ and $Lap(y)$ represent the edges extracted from \hat{y} and y through the Laplacian operator, respectively. The penalty coefficient ε is empirically set to 10^{-3} .

IV. EXPERIMENTS AND DISCUSSION

In this section, to clearly demonstrate the GPANet, the details of the experimental procedure are presented. First, the operating environment and implementation details of network training and learning are presented. Second, we introduce the referenced and non-referenced evaluation indicators used in our experiments. Then, we perform qualitative and quantitative comparisons with traditional and learning-based enhancement methods on real-world paired and non-paired low-light standard test datasets. Fourth, to validate the value of the gradient prior features introduced in this paper, we analyze ablation experiments. Finally, we test the running time of the proposed model on a single high-resolution image.

A. IMPLEMENTATION DETAILS

The imaging environment of low-light images is complex and changeable. To improve the generalization ability and robustness of the deep model, we will adopt two strategies to obtain multi-scene paired training datasets. We propose to train an existing LOL dataset [25] and GLADNet [61]. Specifically, LOL dataset obtains low-/normal-light data pairs by varying exposure parameters in daylight. It contains a total of 1500 pairs of data, of which 500 pairs are collected from real scenes, and the other 1000 pairs are synthetic data. GLADNet dataset contains a total of 5000 pairs of synthetic data.

TABLE 2. Quantitative comparison between our method and state-of-the-arts on the Eval15 from the LOL dataset [25]. The best results are highlighted in red, and the second-best results are highlighted in blue.

Methods	PSNR \uparrow	SSIM \uparrow	FSIM \uparrow	VSI \uparrow	LOE \downarrow	NIQE \downarrow	PIQE \downarrow	AIC \downarrow
HE [4]	14.8006 \pm 3.0900	0.3863 \pm 0.0850	0.8266 \pm 0.0702	0.9220 \pm 0.0349	309.3739 \pm 193.3495	8.4763 \pm 1.1708	15.6537 \pm 3.0869	15.5828 \pm 3.1054
NPE [10]	16.9697 \pm 2.8689	0.4744 \pm 0.1161	0.8964 \pm 0.0312	0.9647 \pm 0.0114	463.5902 \pm 187.5227	8.4390 \pm 1.2128	14.8774 \pm 3.8692	14.8322 \pm 3.9156
SRIE [11]	11.8552 \pm 3.8391	0.4942 \pm 0.1359	0.9085 \pm 0.0424	0.9697 \pm 0.0141	308.7221 \pm 96.3854	7.2869 \pm 1.0779	7.3662 \pm 3.0565	7.4218 \pm 3.0514
LIME [12]	17.1818 \pm 3.8917	0.5558 \pm 0.1129	0.9235 \pm 0.0293	0.9687 \pm 0.0130	300.4954 \pm 60.4202	4.9914 \pm 0.9956	22.5338 \pm 7.9661	23.0535 \pm 8.0154
JIEP [13]	12.0466 \pm 3.9237	0.5104 \pm 0.1335	0.9137 \pm 0.0456	0.9702 \pm 0.0155	287.4193 \pm 89.1650	6.8728 \pm 0.8532	5.4960 \pm 1.7654	5.4983 \pm 1.7602
CRM [21]	17.2033 \pm 4.8653	0.6184 \pm 0.1093	0.9456\pm0.0242	0.9771\pm0.0101	205.0796 \pm 83.3891	7.6865 \pm 1.0424	9.5738 \pm 3.8799	9.5771 \pm 3.8338
Dong [17]	16.7165 \pm 3.9899	0.4722 \pm 0.0991	0.8886 \pm 0.0225	0.9659 \pm 0.0103	347.0275 \pm 148.6002	8.3157 \pm 1.1013	14.9191 \pm 3.0416	14.9355 \pm 3.0141
BIMEF [22]	13.8753 \pm 4.4533	0.5935 \pm 0.1212	0.9263 \pm 0.0316	0.9722 \pm 0.0111	223.7307 \pm 74.6898	7.5150 \pm 1.0890	7.6131 \pm 3.6512	7.6018 \pm 3.6822
DeHz [18]	15.6882 \pm 4.5883	0.5355 \pm 0.1227	0.9311 \pm 0.0273	0.9762 \pm 0.0120	207.6630 \pm 84.6663	7.6368 \pm 1.1550	10.8579 \pm 3.3221	10.9191 \pm 3.3011
RetinexNet [25]	17.1188 \pm 2.5807	0.5961 \pm 0.0864	0.8592 \pm 0.0324	0.9524 \pm 0.0138	424.5449 \pm 257.3204	6.3983 \pm 0.9278	18.4205 \pm 4.5701	18.4272 \pm 4.5757
MBLLEN [26]	17.8583 \pm 3.5495	0.7266 \pm 0.0612	0.9262 \pm 0.0257	0.9698 \pm 0.0067	202.0123\pm85.0360	4.3579 \pm 0.6660	15.0199 \pm 7.2860	15.0680 \pm 6.9888
KinD [27]	17.7129 \pm 3.2958	0.7767\pm0.1014	0.9255 \pm 0.0325	0.9732 \pm 0.0113	405.7201 \pm 235.2285	4.6947 \pm 1.2580	16.1087 \pm 9.9251	15.5222 \pm 8.9528
DeepUPE [55]	12.8956 \pm 4.867	0.4624 \pm 0.1408	0.9009 \pm 0.0541	0.9590 \pm 0.0194	228.0959 \pm 81.0609	7.6368 \pm 1.1550	7.8642 \pm 2.2685	7.8064 \pm 2.1269
EnlightenGAN [29]	17.4475 \pm 4.4921	0.6439 \pm 0.1114	0.9220 \pm 0.0277	0.9701 \pm 0.0135	397.7811 \pm 169.6182	4.7931 \pm 0.8212	9.3802 \pm 3.4784	9.3637 \pm 3.4834
DLN [28]	19.2538\pm4.0512	0.6954 \pm 0.0892	0.9397 \pm 0.0174	0.9749 \pm 0.0098	196.1161\pm67.7769	6.0921 \pm 0.8875	7.5023 \pm 3.1128	7.4929 \pm 3.0534
Zero-DCE [30]	14.8607 \pm 4.2719	0.5588 \pm 0.1246	0.9276 \pm 0.0247	0.9757 \pm 0.0112	274.0245 \pm 86.0038	7.7668 \pm 1.1081	10.3353 \pm 3.8581	10.3369 \pm 3.8694
DeepLPF [64]	11.6887 \pm 3.8703	0.432 \pm 0.1584	0.8569 \pm 0.0749	0.9442 \pm 0.0280	333.2798 \pm 126.5483	3.7092\pm0.6198	9.9836 \pm 5.0784	9.6824 \pm 4.8243
RUAS [33]	16.4047 \pm 4.3879	0.4996 \pm 0.1135	0.8933 \pm 0.0379	0.9569 \pm 0.0168	219.9659 \pm 87.9807	6.3400 \pm 1.1653	3.2627\pm1.1046	3.2748\pm1.1405
DSLR [65]	14.9302 \pm 5.153	0.6059 \pm 0.1305	0.8865 \pm 0.0369	0.9596 \pm 0.0142	376.5956 \pm 136.0030	4.1593 \pm 0.6547	4.2788 \pm 2.6078	4.2445 \pm 2.5754
GPANet	20.862\pm3.7238	0.7842\pm0.1037	0.9400\pm0.0203	0.9780\pm0.0073	219.4380 \pm 105.5755	3.4102\pm0.4891	4.0964\pm1.3747	4.0925\pm1.3697

Furthermore, to further improve the diversity of training data, we randomly rotate and flip the training dataset to enrich the feature structure information of the samples. In the test inference validation, we will perform objective and subjective evaluation analysis on real-world paired and non-paired low-light datasets, including LOL [25], DICM [44], ExDARK [60], LIME [12], MEF [62], MIT-Adobe FiveK [63], NPE [10], and TMDIED¹. Our GPANet is trained and tested on experimental devices with Windows OS, Intel(R) Core(TM) i9-10850K CPU @ 3.60GHz and Nvidia GeForce RTX 2080TI GPU. The framework used for training is Pytorch and the Adam optimizer is used to propose 60 epochs to train GPANet. The initial learning rate is 10^{-3} , and at 20 and 40 epochs, the learning rate is multiplied by 0.1 attenuation. Training time of proposed model can be completed in about 24 hours. For a fair comparison of all traditional and learning-based low-light image enhancement methods, all test codes are downloaded and tested from the code link published in the authors' paper.

B. PERFORMANCE EVALUATIONS

To more comprehensively evaluate the enhanced performance of the proposed model, our GPANet will be compared on synthetic and real-world low-light images with the state-of-the-art 19 current methods, including HE-based methods (i.e., HE [4]), Retinex-based methods (i.e., NPE [10], SRIE [11], LIME [12], and JIEP [13]), camera respond model-based methods (i.e., CRM [21], Dong [17] and BIMEF [22]), dehazing-based methods (i.e., DeHz [18]) and learning-based methods (i.e., RetinexNet [25], MBLLEN [26], KinD [27], DeepUPE [55], EnlightenGAN [29], DLN [28], Zero-DCE [30], DeepLPF [64], RUAS [33], and DSLR [65]). In this subsection, we consider quantitatively analyzing the performance of different enhancement methods from several aspects. Evaluation indicators are roughly divided into two

categories: one is an evaluation with reference, and the other is an evaluation without reference. Reference-based evaluation metrics require the original image to be available, and the enhanced results are compared with the ground truth, while non-reference-based evaluation metrics do not require. Specifically, we will utilize peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [66], feature similarity (FSIM) [67], visual saliency-induced Index (VSI) [68], and lightness order error (LOE) [10] indicators to quantitatively evaluate the enhancement performance under different degradation conditions. The above four metrics evaluate the performance in terms of numerical distance and structural similarity between the augmented result and the corresponding well-exposed image (i.e., the ground truth), respectively. Meanwhile, three popular non-reference image quality assessment methods, including natural image quality evaluator (NIQE) [69], perceptual-based image quality evaluator (PIQE) [70], and AIC (entropy) [71] are used for blind image quality assessment in practical experiments. The above three metrics are based on the entropy of informative regions, the quality-aware set of natural scene statistics models, tone mapping, and the hierarchical perception mechanism (from local structure to global semantics) in human vision systems in a non-referenced manner. These visual quality metrics are expected to provide objective criteria to evaluate the performance of each method from different perspectives.

C. QUANTITATIVE ANALYSIS

1) PAIRD TEST DATASET

To objectively evaluate the performance of our GPANet, we first select the standard test set of 15 paired images from the LOL dataset. According to Table 2, GPANet ranks first in three reference evaluation metrics (i.e., PSNR, SSIM, and VSI) and one no-reference evaluation metric (i.e., NIQE) when compared to the other 19 competing methods. Although our method does not all achieve the best performance in other reference and no-reference evaluations, it still has a

¹<https://sites.google.com/site/vonikakis/datasets/tm-died>

TABLE 3. Quantitative comparison of NIQE between our method and state-of-the-arts on the DICM [30], ExDARK [60], LIME [12], MEF [72], MIT-Adobe FiveK [63], NPE [10], and the TMDIED datasets. The best results are highlighted in red, and the second-best results are highlighted in blue.

Methods	DICM [30]	ExDARK [60]	LIME [12]	MEF [72]	MIT [63]	NPE [10]	TMDIED	Average
NPE [10]	3.760±1.264	3.819±0.959	3.953±1.948	3.529±1.110	2.956±0.856	3.578±1.391	4.665±1.238	3.741±1.108
LIME [12]	4.195±1.355	4.177±0.956	4.869±2.170	4.272±1.073	3.509±1.233	3.912±1.092	3.806±0.860	3.809±1.098
CRM [21]	3.801±1.305	3.855±0.996	3.860±1.712	3.265±1.090	2.952±0.893	3.562±1.059	4.576±1.219	3.709±1.079
DeHz [18]	3.586±1.135	3.770±0.871	3.884±1.788	3.472±1.365	3.135±0.840	3.459±1.074	4.290±1.157	3.649±1.023
KinD [27]	4.123±1.484	3.645±0.861	4.838±3.338	3.845±1.014	3.240±0.899	3.556±0.911	3.395±0.838	3.482±0.972
DLN [28]	3.696±1.183	4.401±1.524	3.868±1.729	3.355±1.098	3.090±0.978	3.786±1.511	4.539±1.230	3.810±1.199
Zero-DCE [30]	3.741±1.287	3.734±0.992	3.742±1.825	3.283±0.999	2.976±0.878	3.661±1.460	4.623±1.197	3.726±1.112
DeepUFE [55]	3.990±1.223	4.233±0.953	4.208±1.410	3.674±0.937	3.067±0.911	3.665±1.100	4.428±1.166	3.780±1.054
DSLR [65]	6.025±0.967	5.730±0.838	5.499±1.772	4.408±0.877	5.115±1.146	4.487±1.352	3.976±0.729	4.814±1.237
GPANet	3.721±1.274	3.266±0.747	3.757±1.596	3.408±0.916	3.101±0.908	3.395±0.847	3.496±0.824	3.349±0.906

TABLE 4. Quantitative comparison of PIQE between our method and state-of-the-arts on the DICM [30], ExDARK [60], LIME [12], MEF [72], MIT-Adobe FiveK [63], NPE [10], and the TMDIED datasets. The best results are highlighted in red, and the second-best results are highlighted in blue.

Methods	DICM [30]	ExDARK [60]	LIME [12]	MEF [72]	MIT [63]	NPE [10]	TMDIED	Average
NPE [10]	15.064±11.945	8.799±2.605	3.931±1.933	8.607±3.384	7.071±2.198	10.274±4.652	8.646±3.592	8.855±3.927
LIME [12]	24.327±15.276	12.136±5.705	4.714±2.049	25.514±11.526	12.210±4.963	17.312±10.405	11.996±6.365	14.130±7.225
CRM [21]	15.290±12.500	9.084±3.023	3.870±1.720	8.096±3.997	7.295±2.360	10.790±5.365	8.133±3.510	8.868±4.143
DeHz [18]	13.528±9.970	8.352±2.382	3.892±1.794	8.496±6.944	5.865±1.972	10.505±5.646	8.122±3.422	8.107±3.786
KinD [27]	21.727±14.126	9.128±3.985	4.710±3.273	19.527±11.413	9.013±3.735	13.531±7.749	8.999±4.900	10.978±5.779
DLN [28]	15.240±13.865	10.348±4.819	3.876±1.735	7.589±4.231	7.556±2.520	10.923±5.721	7.280±3.457	8.799±4.515
Zero-DCE [30]	14.721±11.629	8.713±2.611	3.782±1.860	8.192±3.484	7.233±2.290	10.553±4.895	8.399±3.494	8.816±3.928
DeepUFE [55]	15.301±14.750	5.938±2.422	4.379±1.628	11.413±7.245	6.843±2.512	11.450±6.725	5.517±3.419	7.796±4.574
DSLR [65]	32.886±13.073	30.670±9.388	25.961±18.922	16.624±9.527	21.809±9.975	17.903±12.770	16.317±8.525	21.386±11.714
GPANet	16.215±15.737	6.029±1.282	3.770±1.608	11.170±6.876	6.870±2.014	10.783±6.721	3.496±0.824	7.179±3.581

TABLE 5. Quantitative comparison of AIC between our method and state-of-the-arts on the DICM [30], ExDARK [60], LIME [12], MEF [72], MIT-Adobe FiveK [63], NPE [10], and the TMDIED datasets. The best results are highlighted in red, and the second-best results are highlighted in blue.

Methods	DICM [30]	ExDARK [60]	LIME [12]	MEF [72]	MIT [63]	NPE [10]	TMDIED	Average
NPE [10]	15.087±12.029	8.762±2.603	11.416±8.982	8.588±3.515	7.062±2.200	10.280±4.726	8.620±3.580	8.948±4.042
LIME [12]	24.145±14.995	12.000±5.717	19.521±12.802	25.249±11.057	12.111±5.005	17.283±10.485	11.935±6.321	14.245±7.348
CRM [21]	15.183±12.382	9.079±3.046	11.811±10.101	8.094±3.978	7.292±2.396	10.724±5.320	8.102±3.486	8.950±4.251
DeHz [18]	13.525±9.875	8.395±2.447	11.126±8.881	8.418±6.370	5.889±1.969	10.493±5.573	8.119±3.418	8.216±3.857
KinD [27]	21.873±14.210	9.172±3.915	16.731±11.179	19.394±11.656	8.972±3.572	13.394±7.641	8.943±4.786	11.113±5.793
DLN [28]	15.296±14.021	10.361±4.823	10.593±10.466	7.673±4.188	7.510±2.496	10.864±5.615	7.276±3.427	8.878±4.621
Zero-DCE [30]	14.538±11.396	8.724±2.575	11.172±9.115	8.285±3.698	7.219±2.246	10.541±4.867	8.392±3.482	8.896±3.987
DeepUFE [55]	15.198±14.515	5.863±2.407	14.615±10.050	11.695±7.399	6.877±2.658	11.378±6.665	12.162±5.146	10.003±5.253
DSLR [65]	33.104±13.153	30.934±9.260	24.564±16.797	16.978±10.145	21.886±9.976	17.806±12.589	16.245±8.502	21.411±11.700
GPANet	16.241±15.773	5.984±1.277	12.342±8.250	10.616±6.657	6.820±1.888	10.834±6.675	3.496±0.824	7.273±3.623

high ranking. Compared to some learning-based methods, the performance of traditional methods is relatively stable. On the one hand, it's due to the structure of the deep model's insufficient robustness. In addition, a lack of diversity in the learning data can hinder performance improvement. Our GPANet is able to accurately extract latent gradient features and learn multi-view features at multiple scales. It can enhance learning and reasoning in complex low-light environments.

2) NON-PAIRD TEST DATASET

To further verify the enhanced performance of the proposed method, as shown in Tables 3, 4, and 5, we further analyze the DICM [30], ExDARK [60], LIME [12], MEF [72], MIT-Adobe FiveK [63], NPE [10], and TMDIED dataset standard test datasets through non-reference quantitative evaluation metrics. The above 7 datasets cover a large number of low-light scenes such as traffic, nature, indoors, etc. Learning-based methods have poor robustness and

generalization ability compared to traditional methods, which is a challenge to achieve satisfactory enhancement performance in different complex low-light imaging scenes. Nonetheless, the Retinex-based method is challenging to have better enhancement performance. Camera response model-based methods can adjust the exposure of the image and still have stable enhancement performance on the complex low-light image. Different learning-based methods have obvious differences, which further shows that the generalization ability of deep models is easily limited by the embedded model and learning data. Retinex-based KinD has enhanced performance stability, which verifies that adding additional physical models or prior features to the deep model can improve the robustness of the network. Our GPANet can accurately extract gradient prior features from low-light environments and assist in the restoration of low-light images at multiple scales, which can enhance the processing of images collected in various low-light scenes. Although it



FIGURE 3. Visual comparison of different enhancement methods for three typical images from the LOL dataset [25]. From top-left to bottom-right: (a) Low-light image, restored images, generated by (b) HE [4], (c) NPE [10], (d) SRIE [11], (e) LIME [12], (f) JIEP [13], (g) CRM [21], (h) Dong [17], (i) BIMEF [22], (j) DeHz [18], (k) RetinexNet [25], (l) MBLLEN [26], (m) KinD [27], (n) DeepUPE [55], (o) EnlightenGAN [29], (p) DLN [28], (q) Zero-DCE [30], (r) DeepLPF [64], (s) RUAS [33], (t) DSLR [65], (u) proposed GPANet, and (v) Ground Truth, respectively.

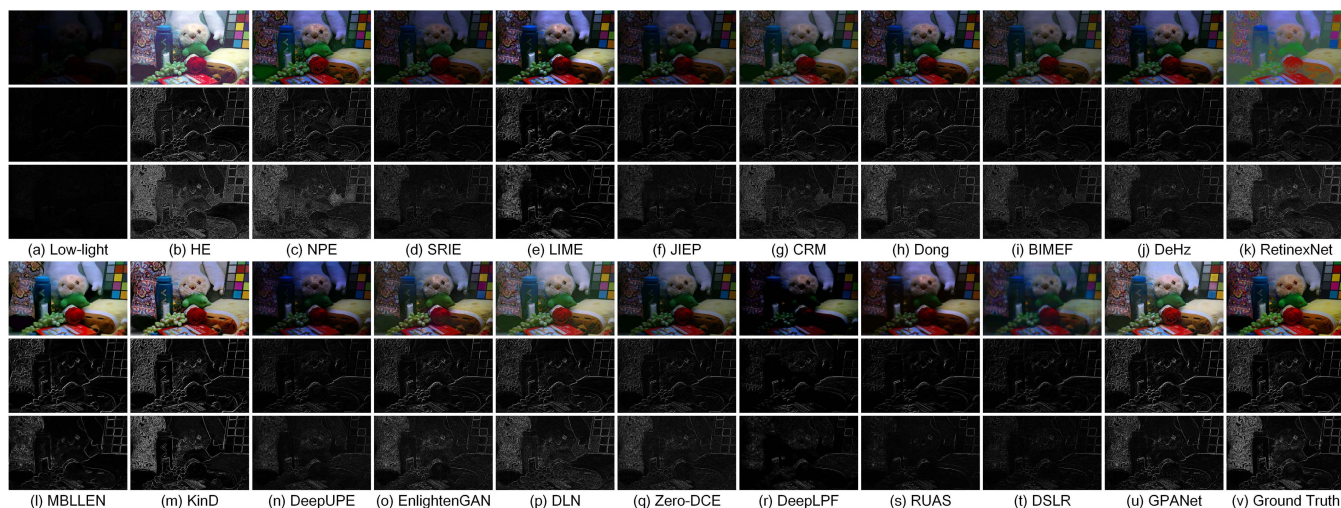


FIGURE 4. Visual comparison of different enhancement methods for one low-light image and its Sobel-based and Laplacian-based gradient features. From top-left to bottom-right: (a) low-light image, restored images, generated by (b) HE [4], (c) NPE [10], (d) SRIE [11], (e) LIME [12], (f) JIEP [13], (g) CRM [21], (h) Dong [17], (i) BIMEF [22], (j) DeHz [18], (k) RetinexNet [25], (l) MBLLEN [26], (m) KinD [27], (n) DeepUPE [55], (o) EnlightenGAN [29], (p) DLN [28], (q) Zero-DCE [30], (r) DeepLPF [64], (s) RUAS [33], (t) DSLR [65], (u) proposed GPANet, and (v) Ground Truth, respectively.

does not achieve the best objective evaluation results in all test datasets, it ranks first in the mean calculation of all test datasets, further validating the robustness and generalizability of GPANet.

D. VISUAL ANALYSIS

To compare the visual performance of start-of-the-art enhancement methods, three real-world low-light images are chosen from the LOL test dataset. As shown in Figure 3, low-light images captured in the real world are largely low in brightness and contrast, accompanied by noisy information that destroys image texture details. HE can enhance the image’s contrast, but there is the phenomenon of color

distortion and noise amplification. NPE can measure image brightness in a more natural manner than other Retinex-based methods (i.e., SRIE, LIME, and JIEP). LIME is enhanced by the denoising function of BM3D, which suppresses unneeded noise points and makes the model appear natural. It is challenging for camera response model-based methods to improve the brightness and contrast of images captured in excessively dark environments. Although RetinexNet decomposes low-light images into illumination components and reflection components, the color of the enhanced images is distorted due to the limitations of its learning data. The images improved by MBLLEN are over-smoothed, the edge texture information is lost locally, and the contrast remains



FIGURE 5. Visual comparison of different enhancement methods for one typical image from the TMDIED dataset. From top-left to bottom-right: (a) Low-light image, restored images, generated by (b) HE [4], (c) NPE [10], (d) SRIE [11], (e) LIME [12], (f) JIEP [13], (g) CRM [21], (h) Dong [17], (i) BIMEF [22], (j) DeHz [18], (k) RetinexNet [25], (l) MBLLEN [26], (m) KinD [27], (n) DeepUPE [55], (o) EnlightenGAN [29], (p) DLN [28], (q) Zero-DCE [30], (r) DeepLPPF [64], (s) RUAS [33], (t) DSLR [65], and (u) proposed GPANet, respectively.

low. There is a slight overexposure phenomenon, but KinD has a strong perception of potential structural features and can improve image contrast while preserving texture structure. DeepUPE, DeepLPPF, and DSLR are challenging to extract valuable information from the dark background when the light intensity is too low. The images enhanced by DLN and Zero-DCE have low contrast, and the color information is lost. Although the color distribution of RUAS is close to the actual value, the texture information is lost, and the image is locally blurred. Our GPANet benefits from the constraints of gradient prior features, which can enhance the contrast and brightness of images without destroying the underlying texture and color features. What’s more, it has a visual performance that is more comparable to that of real

clear images. As shown in Fig. 4, we choose the method with the best performance to compare the edge detection results of Sobel and Laplacian for the enhanced image. Multi-view and multi-scale learning and reasoning enable GPANet to accurately extract gradient information and assist image enhancement in complex low-light imaging environments. To test the robustness of the proposed method on visual evaluation, we randomly select one image from the DICM, ExDARK, LIME, MEF, MIT-Adobe FiveK, NPE, and TMDIED datasets, respectively. Fig. 6 shows the enhancement results with different methods applied to the seven images from the test dataset. Fig. 5 shows the enhancement results on the TMDEID dataset. After zooming in on the local area, it can be clearly found gradient priors-enable GPANet

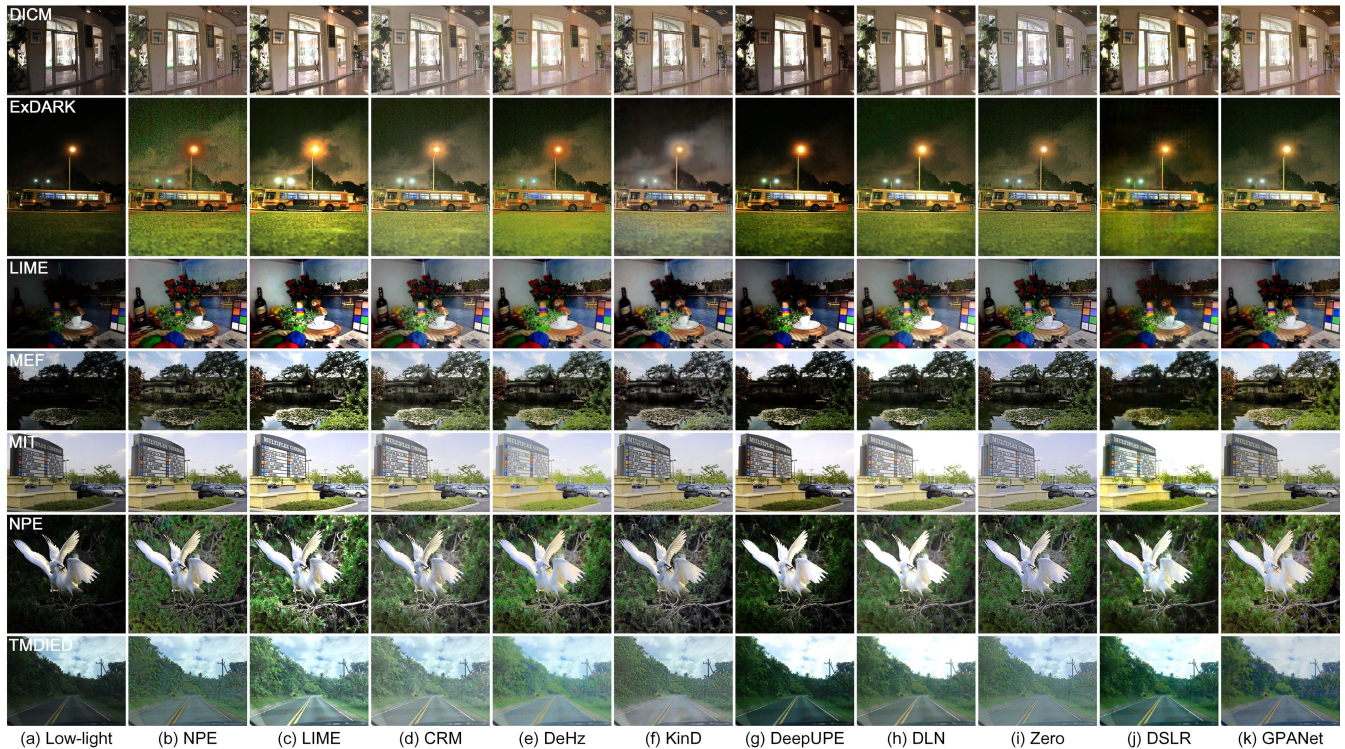


FIGURE 6. Visual comparison of different enhancement methods for seven real-world low-light images in different low-light scenes. From left to right: (a) low-light image, restored images, respectively, generated by (b) NPE [10], (c) LIME [12], (d) CRM [21], (e) DeHz [18], (f) KinD [27], (g) DeepUPE [55], (h) DLN [28], (i) Zero-DCE [30], (j) DSLR [65], and (k) proposed GPANet, respectively.

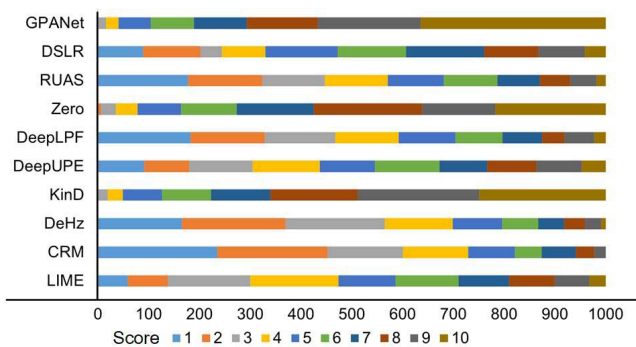


FIGURE 7. Rating distribution of the user study.

can balance image brightness enhancement and edge texture information preservation.

E. USER STUDY

Furthermore, we conduct a user study to understand how our model differs from other state-of-the-art methods, including LIME, CRM, DeHz, KinD, DeepUPE, DeepLPF, Zero-DCE, RUAS, DSLR, and GPANet. The test dataset consists of 20 images from the dataset mentioned by IV-A. For each image in the test method, a score between 1 and 10 is produced. In order to be fair and impartial, each method will be sent to the user in an anonymous way. As shown in Fig. 7, our method has the highest score of 10 and ranks first with

TABLE 6. Ablation experiments on the loss function. The results are shown in PSNR, SSIM, FSIM, VSI, and LOE on the 15 images from the LOL test dataset [25].

	PSNR \uparrow	SSIM \uparrow	FSIM \uparrow	VSI \uparrow	LOE \downarrow
w/o Sobel, w/o Laplacian	19.78	0.767	0.926	0.963	231.62
w Sobel, w/o Laplacian	20.04	0.759	0.921	0.966	229.18
w/o Sobel, w Laplacian	20.15	0.774	0.933	0.971	225.72
w Sobel, w Laplacian	20.86	0.784	0.940	0.978	219.43

an average score of 8.288. Therefore, we can further confirm the performance of the proposed GPANet in low-light image enhancement.

F. ABLATION STUDY

In this section, we attempt to verify the value of gradient prior information in GPANet. We continue to utilize the 15 images from the LOL test dataset for testing ablation. According to the index values provided in Table 6, the objective evaluation performance is lowest when gradient prior information is not included. Adding Sobel or Laplacian prior information improves the performance of a network. When both types of priors were used for deep network learning and inference, PSNR, SSIM, FSIM, VSI, and LOE performance improved by 1.08, 0.017, 0.014, 0.015, and 12.19, respectively, compared to when gradient information was not introduced.

G. OBJECT DETECTION AFTER VISIBILITY ENHANCEMENT

We investigate how high-level vision tasks interact with low-light enhancement by detecting under-lit objects.

TABLE 7. Metric results (mAP) of YOLOv4 on the 4952 voc2007 synthetic low-light images from the enhanced images yielded by enhancement methods.

Classes	bicycle	boat	bus	car	motorbike	person	average
Low-light	0.892	0.754	0.893	0.915	0.901	0.866	0.870
HE [4]	0.885	0.709	0.892	0.907	0.896	0.868	0.860
NPE [10]	0.904	0.753	0.896	0.916	0.897	0.873	0.873
KinD [27]	0.882	0.780	0.870	0.912	0.912	0.871	0.871
GPANet	0.914	0.793	0.900	0.923	0.918	0.887	0.889
Ground Truth	0.931	0.831	0.935	0.941	0.921	0.911	0.912

TABLE 8. Comparisons of the model size and the running time (sec.) for the resolution of 2K image (2560 × 1440 pixels).

Methods	Language	Frame	Model Size	Running Time
HE [4]	Matlab (C)	—	—	0.1850
NPE [10]	Matlab (C)	—	—	70.4131
SRIE [11]	Matlab (C)	—	—	33.3605
LIME [12]	Matlab (C)	—	—	3.2868
JIEP [13]	Matlab (C)	—	—	43.3286
CRM [21]	Matlab (C)	—	—	1.8023
BIMEF [22]	Matlab (C)	—	—	1.6309
Dong [17]	Matlab (C)	—	—	0.9909
DeHz [18]	Python (C)	—	—	51.4564
RetinexNet [25]	Python (G)	Tensorflow	1738kb	1.5160
MBLLEN [26]	Python (G)	Tensorflow	2002kb	0.9910
KinD [27]	Python (G)	Tensorflow	4014kb	0.6540
DeepUPE [55]	Python (G)	Pytorch	2344kb	0.0020
EnlightenGAN [29]	Python (G)	Tensorflow	33774kb	0.3909
DLN [28]	Python (G)	Pytorch	2767kb	0.0125
Zero-DCE [30]	Python (G)	Pytorch	313kb	0.0019
DeepLPIF [32]	Python (G)	Pytorch	6953kb	0.0079
RUSA [33]	Python (G)	Pytorch	31kb	0.0089
DSLRL [56]	Python (G)	Pytorch	175085kb	0.4620
GPANet	Python (G)	Pytorch	16239kb	0.0129

The basic recognition module is the YOLOv4 model [73]. Then we combine the GPANet and Yolov4 models for joint optimization. We use the Pascal-VOC2007 dataset (ground truth) to generate synthetic low-light images. As demonstrated in Table 7, the image enhanced by our GPANet has the most obvious improvement in object detection accuracy. Our GPANet can optimize and enhance the network through auxiliary edge gradient information and has a better ability to extract potential texture features of images. Therefore, the GPANet-driven object detection network has higher detection accuracy than the other three methods.

H. RUNNING TIME

Although modern computers have significantly improved computing power, for most algorithms, real-time processing of videos with high resolution is still a severe challenge. This section compares traditional and learning methods' running time on the same computers mentioned with Windows OS, Intel(R) Core(TM) i9-10850K CPU @3.60GHz, and Nvidia GeForce RTX 2080TI GPU. As shown in Table 8, we perform low light enhancement on a 2k image (2560 × 1440). The original HE algorithm needs to count the information of the entire image. Though the required computing time is short, it is insufficient to meet real-time processing needs. In general, Retinex-based methods take a longer time. Although the learning methods make use of a powerful GPU, methods other than Zero-DCE still require additional computing

time for mapping. In comparison to other methods, our GPANet is capable of meeting the requirements for real-time enhancement of low-light videos while also providing superior enhancement effects.

V. CONCLUSION AND DISCUSSION

This paper proposes a gradient prior-aided enhancement solution and implements it by introducing first-order (i.e., Sobel Filter) and second-order (i.e., Laplacian Filter) gradient prior features to handle low-light image enhancement. The key is to construct an end-to-end network consisting of an encoder, a multi-branch topology module, and three sub-decoders. GPANet is able to improve the enhancement performance by fusing and decomposing multi-view and multi-scale features. Extensive experiment results demonstrate the advantages of GPANet compared to other methods from both qualitative and quantitative perspectives in maritime-related and other natural low-light scenes. To make our work more reliable and applicable, the research shown in this work can be extended in the following directions.

- To further improve the robustness and effectiveness of our GPANet, the size of the loss function weights cannot rely solely on experimental experience. The weight of the loss function should be adaptively adjusted and updated for different brightness distributions, scene compositions, etc. To this end, we will consider using an equatorial uncertainty weighting strategy to adaptively adjust the constant weights.
- Sobel and Laplacian are the basic first- and second-order gradient operators, and the gradient information generated by them has limitations such as noise interference and insensitivity to grayscale changes. In order to extract edge gradient features from the dark background more accurately, we will try different gradient detection operators or algorithms such as Laplacian of Gaussian, Prewitt, Canny, etc., which can contribute to balance edge extraction and brightness enhancement.
- The generalization ability of deep networks in different low-light scenarios remains a challenge. Therefore, how to adapt the network to different low-light scenarios is one of the urgent problems to be solved. It is a crucial way to improve the stability of the network for complex low-light imaging by introducing physical prior knowledge to constrain the brightness, contrast, and structure of the generated image. We will explore more physical priors and fusion methods of deep networks to come up with better solutions for low-light image enhancement.

REFERENCES

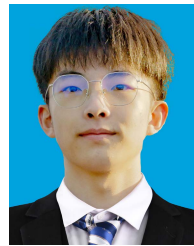
- [1] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image-adaptive Yolo for object detection in adverse weather conditions," 2021, *arXiv:2112.08088*.
- [2] R. W. Liu, W. Yuan, X. Chen, and Y. Lu, "An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system," *Ocean Eng.*, vol. 235, Sep. 2021, Art. no. 109435.

- [3] F. Lv, Y. Li, and F. Lu, "Attention guided low-light image enhancement with a large scale low-light simulation dataset," *Int. J. Comput. Vis.*, vol. 129, no. 7, pp. 2175–2193, Jul. 2021.
- [4] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. T. H. Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Comput. Vis., Graph., Image Process.*, vol. 39, no. 3, pp. 355–368, 1987.
- [5] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Müller, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," in *Proc. 1st Conf. Vis. Biomed. Comput.*, 1990, pp. 337–338.
- [6] E. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Müller, M. P. Braeuning, and S. M. Pizer, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *J. Digit. Imag.*, vol. 11, no. 4, pp. 193–200, Nov. 1998.
- [7] S. F. Tan and N. A. M. Isa, "Exposure based multi-histogram equalization contrast enhancement for non-uniform illumination images," *IEEE Access*, vol. 7, pp. 70842–70861, 2019.
- [8] E. H. Land and J. J. McCann, "Lightness and Retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, 1971.
- [9] E. H. Land, "The Retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, 1977.
- [10] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.
- [11] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2782–2790.
- [12] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2016.
- [13] B. Cai, X. Xu, K. Guo, K. Jia, B. Hu, and D. Tao, "A joint intrinsic-extrinsic prior model for Retinex," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4000–4009.
- [14] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust Retinex model," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.
- [15] Q. Zhang, Y. Nie, and W.-S. Zheng, "Dual illumination estimation for robust exposure correction," in *Computer Graphics Forum*, vol. 38, no. 7, Hoboken, NJ, USA: Wiley, Nov. 2019, pp. 243–252.
- [16] Q. Pan, L. Zhao, S. Chen, and X. Li, "Fusion of low-quality visible and infrared images based on multi-level latent low-rank representation joint with Retinex enhancement and multi-visual weight information," *IEEE Access*, vol. 10, pp. 2140–2153, 2021.
- [17] X. Dong, G. Wang, Y. Pang, W. Li, J. Wen, W. Meng, and Y. Lu, "Fast efficient algorithm for enhancement of low lighting video," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2011, pp. 1–6.
- [18] X. Jiang, H. Yao, S. Zhang, X. Lu, and W. Zeng, "Night video enhancement using improved dark channel prior," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 553–557.
- [19] L. Li, R. Wang, W. Wang, and W. Gao, "A low-light image enhancement method for both denoising and contrast enlarging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3730–3734.
- [20] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Sep. 2010.
- [21] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A new low-light image enhancement algorithm using camera response model," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 3015–3022.
- [22] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," 2017, *arXiv:1711.00591*.
- [23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, Sep. 2015.
- [24] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.
- [25] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep Retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf.*, 2018, pp. 1–12.
- [26] F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: Low-light image/video enhancement using CNNs," in *Proc. Brit. Mach. Vis. Conf.*, 2018, p. 220.
- [27] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. ACM Int. Conf. Multimedia*, 2019, pp. 1632–1640.
- [28] L.-W. Wang, Z.-S. Liu, W.-C. Siu, and D. P. Lun, "Lightening network for low-light image enhancement," *IEEE Trans. Image Process.*, vol. 29, pp. 7984–7996, 2020.
- [29] Y. Jiang, X. Gong, D. Liu, Y. Cheng, and C. Fang, "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.
- [30] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1780–1789.
- [31] Y. Guo, Y. Lu, R. W. Liu, M. Yang, and K. T. Chui, "Low-light image enhancement with regularized illumination optimization and deep noise suppression," *IEEE Access*, vol. 8, pp. 145297–145315, 2020.
- [32] F. Zhang, Y. Li, S. You, and Y. Fu, "Learning temporal consistency for low light video enhancement from single images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 4967–4976.
- [33] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10561–10570.
- [34] R. Wang, B. Jiang, C. Yang, Q. Li, and B. Zhang, "MAGAN: Unsupervised low-light image enhancement guided by mixed-attention," *Big Data Mining Anal.*, vol. 5, no. 2, pp. 110–119, Jun. 2022.
- [35] R. Al Sobhahi and J. Tekli, "Low-light homomorphic filtering network for integrating image enhancement and classification," *Signal Process., Image Commun.*, vol. 100, Jan. 2022, Art. no. 116527.
- [36] Y. Fu, Y. Hong, L. Chen, and S. You, "LE-GAN: Unsupervised low-light image enhancement network using attention module and identity invariant loss," *Knowl.-Based Syst.*, vol. 240, Mar. 2022, Art. no. 108010.
- [37] I. E. Sobel, *Camera Models and Machine Perception*. Stanford, CA, USA: Stanford University, 1970.
- [38] X. Wang, "Laplacian operator-based edge detectors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 886–890, May 2007.
- [39] X. Zhang, H. Zeng, and L. Zhang, "Edge-oriented convolution block for real-time super resolution on mobile devices," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 4034–4043.
- [40] Y. Liu, S. Anwar, L. Zheng, and Q. Tian, "GradNet image denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 508–509.
- [41] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 593–600, May 2007.
- [42] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *J. VLSI Signal Process. Syst. Signal, Image Video Technol.*, vol. 38, no. 1, pp. 35–44, 2004.
- [43] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1752–1758, Nov. 2007.
- [44] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2D histograms," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5372–5384, Dec. 2013.
- [45] S. Kansal, S. Purwar, and R. K. Tripathi, "Image contrast enhancement using unsharp masking and histogram equalization," *Multimedia Tools Appl.*, vol. 77, no. 20, pp. 26919–26938, 2018.
- [46] N. H. Saad, N. A. M. Isa, and H. M. Saleh, "Nonlinear exposure intensity based modification histogram equalization for non-uniform illumination image enhancement," *IEEE Access*, vol. 9, pp. 93033–93061, 2021.
- [47] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "Properties and performance of a center/surround Retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.
- [48] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "A multiscale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.
- [49] Z.-U. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale Retinex for color image enhancement," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Sep. 1996, pp. 1003–1006.
- [50] X. Ren, W. Yang, W.-H. Cheng, and J. Liu, "LR3M: Robust low-light enhancement via low-rank regularized Retinex model," *IEEE Trans. Image Process.*, vol. 29, pp. 5862–5876, 2020.

- [51] W. Wang, Z. Chen, and X. Yuan, "Simple low-light image enhancement based on Weber–Fechner law in logarithmic space," *Signal Process., Image Commun.*, vol. 106, Aug. 2022, Art. no. 116742.
- [52] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [53] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–12, Jul. 2017.
- [54] C. Li, J. Guo, F. Porikli, and Y. Pang, "LightenNet: A convolutional neural network for weakly illuminated image enhancement," *Pattern Recognit. Lett.*, vol. 104, pp. 15–22, Mar. 2018.
- [55] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 6849–6857.
- [56] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang, "Beyond brightening low-light images," *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 1013–1037, 2021.
- [57] Z. Zhao, B. Xiong, L. Wang, Q. Ou, L. Yu, and F. Kuang, "RetinexDIP: A unified deep framework for low-light image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1076–1088, Mar. 2022.
- [58] A. Garg, X.-W. Pan, and L.-R. Dung, "LiCENet: Low-light image enhancement using the light channel of HSL," *IEEE Access*, vol. 10, pp. 33547–33560, 2022.
- [59] Y. Lu, Y. Guo, R. W. Liu, and W. Ren, "MTRBNet: Multi-branch topology residual block-based network for low-light enhancement," *IEEE Signal Process. Lett.*, vol. 29, pp. 1127–1131, 2022.
- [60] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," *Comput. Vis. Image Understand.*, vol. 178, pp. 30–42, Jan. 2019.
- [61] W. Wang, C. Wei, W. Yang, and J. Liu, "GladNet: Low-light enhancement network with global awareness," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 751–755.
- [62] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
- [63] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *Proc. CVPR*, 2011, pp. 97–104.
- [64] S. Moran, P. Marza, S. McDonagh, S. Parisot, and G. Slabaugh, "DeepLPP: Deep local parametric filters for image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12826–12835.
- [65] S. Lim and W. Kim, "DSLR: Deep stacked Laplacian restorer for low-light image enhancement," *IEEE Trans. Multimedia*, vol. 23, pp. 4272–4284, 2020.
- [66] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [67] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [68] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Aug. 2014.
- [69] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.
- [70] N. Venkatanath, D. Praneth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *Proc. 21st Nat. Conf. Commun. (NCC)*, Feb. 2015, pp. 1–6.
- [71] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
- [72] C. Lee, C. Lee, Y.-Y. Lee, and C.-S. Kim, "Power-constrained contrast enhancement for emissive displays based on histogram equalization," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 80–93, Jan. 2012.
- [73] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.



YUXU LU received the B.Sc. degree in navigation technology from the School of Navigation, Wuhan University of Technology, Wuhan, China, in 2020, where he will pursue the M.S. degree in traffic information engineering and control with the School of Navigation. He is currently working as a Research Assistant with the Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong. His research interests include computer vision, machine learning, and intelligent navigation systems. He has received the Excellent Oral Presentation from the 11th International Conference on Machine Learning and Computing.



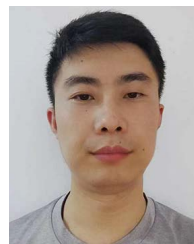
YUAN GAO is currently pursuing the degree with the School of Navigation, Wuhan University of Technology, Wuhan, China. His research interests include computer vision, machine learning, and intelligent navigation systems. He was awarded the National Scholarship by the Ministry of Education of the People's Republic of China, in 2021. His previous work on deep learning-based joint contrast enhancement and noise suppression of low-light images has been accepted by the 6th International Conference on Robotics and Automation Sciences.



YONGQI GUO received the M.S. degree in computational mathematics from the Department of Mathematics, Wuhan University of Technology, Wuhan, China, in 2010. He is currently an Associate Researcher with the Wuhan University of Technology. His research interests include computer vision and mathematical modeling.



WENYU XU received the B.Sc. degree in information engineering from the School of Mechanical Electronic and Information Engineering, China University of Mining and Technology, Beijing, China, in 2020. She is currently working as an Information Engineer with Wuhan Baosight Software Company Ltd., Wuhan, China. Her research interests include computer vision and machine learning.



XIANJUN HU received the Ph.D. degree in system engineering from the Naval University of Engineering, Wuhan, China, in 2013. He is currently a Senior Lecturer with the Naval University of Engineering. His research interests include system optimization, artificial intelligence, and remote sensing applications in coastal and lake ecosystems.

...