



# OPEN Bilateral enhancement network with signal-to-noise ratio fusion for lightweight generalizable low-light image enhancement

Junfeng Wang<sup>1</sup>, Shenghui Huang<sup>1</sup>, Zhanqiang Huo<sup>1✉</sup>, Shan Zhao<sup>1</sup> & Yingxu Qiao<sup>2</sup>

Low-light image enhancement aims to enhance the visibility and contrast of low-light images while eliminating complex degradation issues such as noise, artifacts, and color distortions. Most existing low-light image enhancement methods either focus on quality while neglecting computational efficiency or have limited learning and generalization capabilities. To address these issues, we propose a Bilateral Enhancement Network with signal-to-noise ratio fusion, called BiEnNet, for lightweight and generalizable low-light image enhancement. Specifically, we design a lightweight Bilateral enhancement module with SNR (Signal-to-Noise Ratio) Fusion (BSF), which serves the SNR map of the input low-light image as the interpolation weights to dynamically fuse global brightness features and local detail features extracted from a bilateral network and achieve differentiated enhancement across different regions. To improve the network's generalization ability, we propose a Luminance Normalization (LNM) module for preprocessing and a Dual-Exposure Processing (DEP) module for post-processing. LNM divides the channels of input features into luminance-related channels and luminance independent channels, and reduces the inconsistency of the degradation distribution of input low-light images by only normalizing the luminance-related channels. DEP learns overexposure and underexposure corrections simultaneously by employing the ReLU activation function, inverting operation, and residual network, which can improve the robustness of enhancement effects under different exposure conditions while reducing network parameters. Experiments on the LOL-V1 dataset shows BiEnNet significantly increased PSNR by 8.6% and SSIM by 3.6% compared to FLW-Net, reduced parameters by 98.78%, and improved computational speed by 52.64% compared to the classical KIND.

**Keyword** Low-light Image Enhancement, Lightweight, Generalization ability, SNR Fusion, Normalization

Due to the influence of factors such as shooting environment and equipment limitations, images captured in low-light environments often exhibit various issues, including low brightness, low contrast, severe noise, and uneven color distribution. These image quality problems not only impair visual perception but also significantly impact subsequent downstream computer vision tasks, such as semantic segmentation<sup>1,2</sup> and object detection<sup>3,4</sup>. In recent years, numerous low-light image enhancement methods<sup>5–14</sup> have been proposed. Although these methods have achieved impressive enhancement results, striking a balance between efficiency and quality remains challenging.

Low-light image enhancement methods can be broadly categorized into two categories: traditional methods (e.g., histogram equalization<sup>15–17</sup>, Retinex model<sup>18–22</sup>) and deep learning methods (e.g., MBLLEN<sup>8</sup>, SNR-Aware<sup>23</sup>, and SKF<sup>24</sup>). The evolution from traditional methods to deep learning approaches marks a significant advancement in low-light image enhancement. Traditional methods rely on physical modeling and optimization of image degradation, using hand-crafted algorithms to achieve enhancement. However, as data availability and computational power have grown, deep learning approaches have emerged, leveraging neural networks to learn mappings from input to output, resulting in more precise and efficient low-light enhancement.

<sup>1</sup>School of Software, Henan Polytechnic University, Jiaozuo 454000, China. <sup>2</sup>School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454003, China. ✉email: hzq@hpu.edu.cn

Traditional methods, such as CLAHE<sup>15</sup>, improve the detection of fine structures in mammographic images through contrast-limited adaptive histogram equalization. However, these methods often encounter challenges in complex scenes, such as over-enhancement or noise amplification. Moreover, they require substantial manual prior information for parameter tuning, increasing complexity and limiting their flexibility and applicability in real-world scenarios.

Deep learning methods address some of these issues by training neural networks on large datasets, enabling automatic learning of the mapping from low-light to enhanced images. These methods offer notable improvements in accuracy, robustness, and speed. Nevertheless, they also have limitations. For instance, SKF<sup>24</sup> enhances low-light images using semantic-aware guidance but suffers from a complex network structure and large model parameters, leading to prolonged processing times and lower computational efficiency. Furthermore, because deep learning models inherently learn mappings between input and output domains, variations among samples make them heavily reliant on training data, reducing their generalization capabilities.

In this paper, we opt to normalize the degradation before low-light image enhancement to make the input images have a more consistent degradation distribution. For this, we designed a lightweight Luminance Normalization (LNM) module to normalize the luminance-related channels. The LNM consists of a normalization module for processing luminance information and a gating module for channel selection. Initially, the normalization module normalizes each channel separately, then the gating module filters out the luminance-related channels, and finally, the normalized channels are merged with the original channels. This method can reduce the differences between samples while reducing the loss of information due to normalization and improving the generalization of the model. Considering the variability of exposure conditions, we use a simple network to simultaneously learn the correction of two exposure attribute features. To achieve this, we design the Dual-Exposure Processing (DEP) module, which primarily comprises an exposure activation module and an exposure learning module. Initially, the exposure activation module extracts underexposed and overexposed features. Then, the exposure learning module concurrently learns to correct these features. Finally, we fuse the processed features to enhance the network's robustness across various exposure conditions.

For the enhancement part of the network, considering that different regions of low-light images have varying degrees of brightness and noise degradation conditions, regions with very low brightness have more noise and cannot achieve effective enhancement by relying solely on local information. Conversely, regions with higher brightness achieve good enhancement using only local region information. Therefore, our solution for the enhancement part is to utilize both global and local features. To this end, we design the Bilateral Enhancement module with SNR Fusion (BSF). The global branch, taking the luminance channel and its histogram of the low-light image as inputs, captures global information. Meanwhile, the local branch, with a residual connection structure, captures local information. Then, guided by the SNR prior information of low-light images, it dynamically fuses global and local features to achieve better low-light image enhancement.

In our extensive experiments conducted on representative datasets (LOL-V1<sup>25</sup>, LOL-V2<sup>26</sup>), as well as a mixed dataset, the results show that our BiEnNet recovers more realistic color tones and better contrast and detail compared to other methods (see Fig. 1). Overall, our work makes the following key contributions:



**Fig. 1.** Visual comparison with some other state-of-the-art methods. It can be observed that the enhanced images generated by our method closely resemble the reference images in terms of color and details.

- We propose an LNM module that selects luminance-related channels for normalization, thus enhancing the network's generalization ability under unknown conditions. It is lightweight and can be easily integrated into other tasks.
- To further improve the network's robustness, we devise a DEP module, which simultaneously learns both underexposure and overexposure features within a single network. It enhances the network's ability to handle exposure variations, and like the LNM module, it is lightweight and adaptable for integration into other tasks.
- We design a lightweight BSF module, which is a dual-branch module. The two branches capture global and local features, respectively. By dynamically fusing these features based on SNR priors, we achieve better low-light image enhancement.

## Related work

### Low-light image enhancement

Researchers have focused on low-light enhancement for many years, mainly dividing it into two categories: traditional methods and deep learning methods. Among them, traditional methods mainly include histogram equalization<sup>15–17</sup> and the Retinex model<sup>18–22</sup>. Histogram equalization adjusts overall brightness by expanding the grayscale distribution of an image. Zuiderveld et al.<sup>16</sup> proposed local region histogram equalization, which effectively reduces noise amplification by limiting the upper bound of contrast enhancement. Lee et al.<sup>17</sup> introduced using a tree-structured hierarchical 2D histogram to represent grayscale differences in high-frequency regions. Based on color constancy, the Retinex theory decomposes the original image into an illumination map and reflectance map. Fu et al.<sup>20</sup> utilized a weighted variational model to preserve more detailed reflectance. Li et al.<sup>21</sup> improved the performance of low-light enhancement by introducing noise mapping into the Retinex model. Although these methods have achieved excellent results in enhancing brightness and contrast, they still have significant limitations, such as unsatisfactory noise removal and color restoration.

With the rapid development of deep learning in the field of computer vision, these techniques have been successfully applied to the low-light enhancement field and have become mainstream methods. Lv et al.<sup>8</sup> proposed a multi-branch enhancement network capable of extracting features at different levels and fusing them to generate output images. Jiang et al.<sup>27</sup> introduced the first unsupervised low-light image enhancement method, enabling training without paired data. Guo et al.<sup>6</sup> proposed Zero-DCE, which designs deep networks to estimate dynamically adjustable pixels and curves to achieve brightness enhancement. The structure of URetinex-Net proposed by Wu et al.<sup>28</sup> consists of initialization, optimization, and illumination adjustment modules, which achieve noise suppression and detail preservation. Compared with traditional methods, deep learning methods can learn complex feature representations from massive data, resulting in clearer and more naturally enhanced results. However, due to these typically involving large-scale network structures, they require substantial computational resources during both training and deployment, resulting in longer processing times and making them unsuitable for real-time response applications.

### Model generalization

The generalization ability of a model refers to its performance on unseen datasets, specifically whether it can successfully transfer and apply the knowledge learned from training data to other scenarios. The strength of generalization ability is an important criterion for measuring whether a model has practical application value. Therefore, how to train models with better generalization ability using limited datasets is one of the important topics in deep learning research. In early machine learning algorithms, researchers proposed many methods to address this issue, such as regularization<sup>29,30</sup> and cross-validation<sup>31</sup>.

However, with the increasing complexity and scale of models, previous generalization methods are no longer sufficient to meet practical application needs. Researchers have developed various methods to address this challenge, including domain generalization<sup>32</sup>, transfer learning<sup>33</sup>, meta-learning<sup>34</sup>, zero-shot learning<sup>35</sup>, self-supervised learning<sup>6,21,36</sup>, and adversarial training<sup>37</sup>. Current methods mainly address this issue from the perspectives of datasets or optimization algorithms. Single-domain generalization<sup>38</sup> has also recently gained attention, aiming to train models from a single source domain that can generalize well to other unseen domains. However, these methods typically involve complex network structures, making them unsuitable for real-time application needs. They also do not solve the problem from the perspective of consistency in the degradation distribution of the input images.

### Exposure correction

In the field of digital image processing, exposure correction is a crucial aspect. Traditional exposure correction methods mainly include histogram equalization<sup>39</sup>, gamma correction<sup>40</sup>, and the Retinex model<sup>18,20</sup>. Reza et al.<sup>39</sup> proposed CLAHE, which corrects exposure by adaptively adjusting the histograms of different regions. LIME<sup>18</sup> uses the maximum intensity of the RGB channels as an initial rough illumination map and then refines it using prior structures. However, because these methods heavily rely on manual design and neglect the relationships between pixels, they often exhibit unnatural results.

In recent years, deep learning-based exposure correction methods have gradually emerged<sup>41–45</sup>. Mertens et al.<sup>41</sup> proposed a method that suggests blending well-exposed regions from a sequence of images with different exposure levels into a single high-quality image. However, it requires a multi-exposure image sequence as input, so it cannot apply directly to a single image. Zhang et al.<sup>42</sup> first used sampled tone mapping curves to construct multi-exposure image sequences for each video frame. Then, they gradually fuse the image sequences in a spatiotemporal manner to obtain enhanced videos, thus applying the technique to exposure-deficient video enhancement. More recently, Afifi et al.<sup>43</sup> developed a pyramid-based network to correct exposure in a coarse-to-fine manner, initially restoring brightness and subsequently refining details. Although these methods have



achieved good results, they do not fully utilize both overexposure and underexposure features simultaneously, resulting in suboptimal correction effects.

## Methods

As shown in Fig. 2, BiEnNet primarily consists of three primary components: the Luminance Normalization (LNM) module, the Bilateral Enhancement module with SNR Fusion (BSF), and the Dual-Exposure Processing (DEP) module. Given a low-light image  $I_L$ , the network first expands the channel dimension of the low-light image using the encoder. Then, the LNM module normalizes the channels related to the luminance information of the features, and a decoder obtains the normalized low-light image  $I_N$  to provide a low-light image with a more consistent luminance distribution for the subsequent enhancement. For the low-light enhancement part, the BSF module obtains global luminance features and local detail features of  $I_N$  through the Global Brightness Adjustment (GBA) module and the Local Feature Extraction (LFE) module, respectively. Concurrently, the SNF module employs a non-learning method to obtain an SNR map of  $I_N$ , which guides the dynamic fusion of global and local features. Finally, the DEP module's role is to learn feature representations under different exposure conditions simultaneously, enhancing the network's robustness to various exposure conditions.

### Luminance normalization module

#### Motivation

Due to different lighting conditions and camera parameters, the actual images obtained often exhibit different levels of degradation. This inconsistency among samples poses a challenge for a well-trained model, especially for images with degraded conditions that are not present in the training data. A common approach is to increase the diversity of the training dataset to expand its capacity. However, the high cost of data collection often makes this method impractical. Moreover, a more diverse dataset may increase the difficulty of model training, potentially leading to instability in the training process.

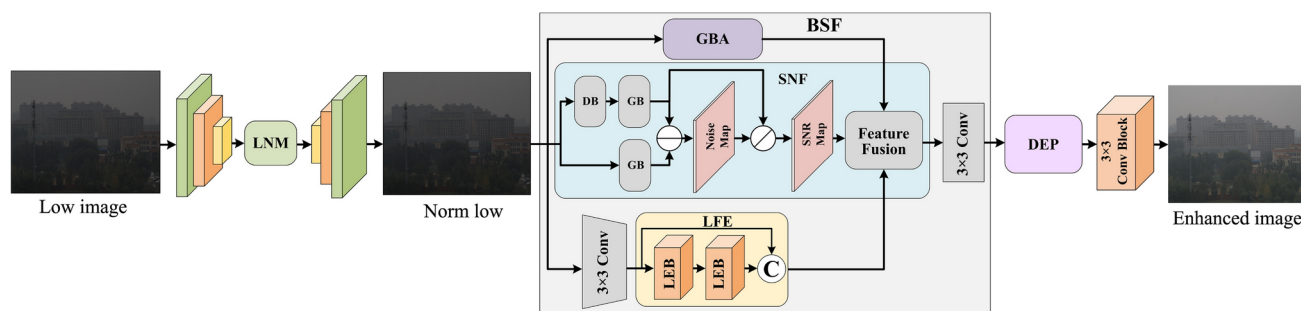
Normalization possesses the capability of reducing differences in image brightness, allowing for a more consistent brightness distribution across different images. This assists the model in effectively extracting information that is not related to brightness, mitigating the impact of brightness variations on model training, reducing the difficulty of subsequent operations, and improving the model's generalization capabilities. Therefore, we choose to apply normalization methods to images with different brightness levels to achieve a more consistent brightness distribution and improve the model's generalization performance.

#### Luminance normalization

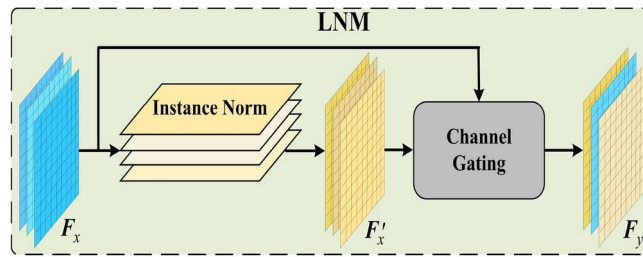
As shown in Fig. 3, the LNM mainly consists of two parts: a normalization module that processes brightness information and a gating module for channel selection. Since Instance Normalization (IN)<sup>46</sup> is unaffected by the number of channels and batch size, and its computation is relatively simple, we use IN for channel normalization. Assume the input feature map  $F_x \in \mathbb{R}^{C \times H \times W}$ , where  $C$  is the number of channels, and  $H$  and  $W$  are the height and width of the feature map, respectively. For each channel feature in  $F_x$ , IN first calculates its mean  $\mu(F_{x,c})$  and variance  $\sigma(F_{x,c})$ , then subtracts the mean  $\mu(F_{x,c})$  and divides by the variance  $\sigma(F_{x,c})$ , and finally scales and shifts the result. We can represent this process as:

$$F'_x = \bigcup_c \text{IN}(F_{x,c}) = \bigcup_c \left( \alpha_c \frac{F_{x,c} - \mu(F_{x,c})}{\sqrt{\sigma^2(F_{x,c}) + \vartheta}} + \eta_c \right) \quad (1)$$

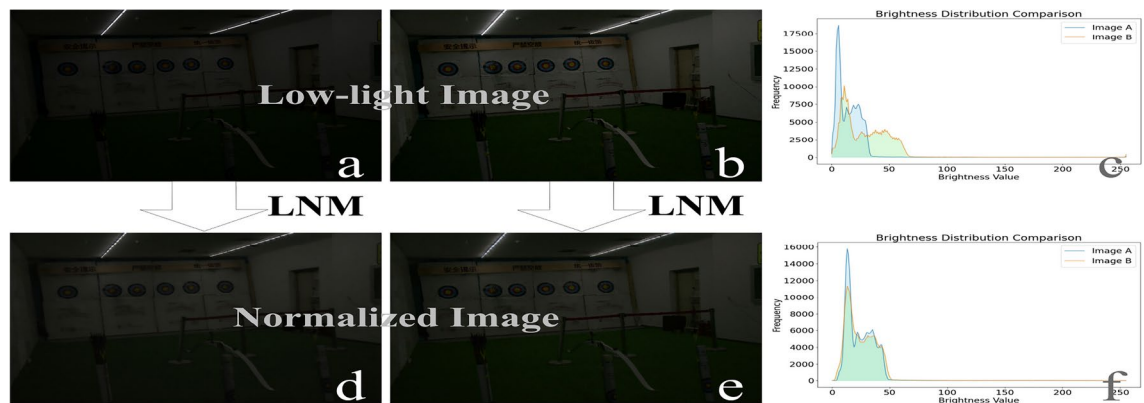
where  $\mu(F_{x,c})$  and  $\sigma(F_{x,c})$  are computed for each channel,  $\bigcup_c$  represents the merging of all normalized channels, and  $\alpha_c$  and  $\eta_c$  are learnable scaling and shifting parameters, and  $\vartheta$  is a very small constant used to prevent division by zero. This method ensures that each channel of every sample has its own mean and variance, thus maintaining the independence between samples. Therefore, this approach can effectively reduce the brightness differences between samples, thereby improving the model's generalization ability.



**Fig. 2.** Detailed architecture of BiEnNet network. Before low-light enhancement, we use LNM to normalize brightness degradation, reducing inconsistencies in images. During enhancement, BSF employs signal-to-noise ratio fusion, using SNF to compute the SNR map  $S$  for low-light images, with DB and GB representing denoising and grayscale computation. GBA and LFE capture global brightness and local detail features, respectively, and then BSF dynamically fuses them through  $S$ . Finally, DEP learns the correction for two exposure conditions, enhancing network robustness.



**Fig. 3.** Detailed structure of the brightness normalization (LNM) module. LNM consists mainly of a per-channel IN normalization layer composed of multiple INs and a gating module for selecting luminance-relevant channels.



**Fig. 4.** Comparison of brightness distributions between low-light images and LNM-normalized results. The low-light images are from the LOL-V2<sup>26</sup> test dataset. The images on the left have the same content but different brightness distributions; the images on the right show the brightness distributions before and after LNM. The results show that the normalized images have similar brightness distributions after applying LNM.

Although the benefits of normalization in reducing sample variations and enhancing model stability, it inevitably leads to some information loss. For instance, it can affect the correlation between channels, potentially impacting the model's accuracy to some extent. Therefore, to mitigate the information loss caused by normalization, we introduce a gating mechanism for channel selection. We expect the gating module to output values close to 0 or 1 to control which channels require normalization. This is specifically expressed as follows:

$$F_y = (1 - G) \odot F_x + G \odot F'_x \quad (2)$$

where,  $G$  represents the gating module, which outputs values of 0 or 1. And  $\odot$  denotes channel-wise multiplication, which is used to selectively weights the normalized features  $F'_x$  and the original features  $F_x$  based on the gating weights  $G$ . We expect  $G$  to dynamically output 0 or 1 according to different features, thereby selecting the channels that truly require normalization. The Sigmoid activation function maps input values to the range  $[0, 1]$ , making it suitable for representing probabilities or weights for normalizing data. Inspired by this property, we design  $G$  as:

$$G = \frac{\beta^2}{\beta^2 + \varepsilon} \quad (3)$$

where  $\beta$  is the output vector of feature  $F_x$  through the convolution layer with activation function, and  $\varepsilon$  is a very small constant to prevent division by zero. We control the value of  $G$  using  $\beta$ , enabling  $G$  to filter the channel effectively.

When  $\beta = 0$ , the normalized image is the same as the original low-light image, i.e.,  $F_y = F_x$ . And when  $\beta \neq 0$ , then  $G \approx 1$  and  $F_y = F'_x$ . Since  $\beta$  is generated by convolution operations,  $G$  easily outputs as 1. To prevent the normalized image from being identical to the original image, we set the normalization operation as in Eq. 2, making  $G$  more inclined towards normalized channels. As shown in Fig. 4, we plot the brightness distribution of images with the same content but different brightness distributions. After processing with the trained LNM, they exhibit similar brightness distributions, further demonstrating the effectiveness of our LNM module.

### Bilateral enhancement module with SNR fusion

Low-light images exhibit varying characteristics such as brightness and noise across different regions. In the same low-light image, regions with lower brightness suffer more severe noise degradation, while regions with higher brightness experience less damage, resulting in relatively better visibility. Most existing methods primarily focus on capturing global information but overlook the imbalance in characteristics across different regions. This may lead to insufficient enhancement in lower brightness regions and over-enhancement in higher brightness regions.

For low-light regions heavily affected by noise, local features alone cannot achieve effective enhancement due to the limited amount of useful information available. In contrast, regions with less noise degradation can be effectively enhanced using only local features.

To address this issue, we employ a dynamic enhancement strategy to enhance pixels in different regions. For regions with high SNR, we enhance them primarily through local information, as they contain sufficient useful information. For regions with low SNR, where noise severely affects local information and useful details are scarce, we utilize global information to enhance them effectively.

Based on this idea, we propose a Bilateral Enhancement module with SNR Fusion (BSF), as shown in Fig. 2; it mainly consists of three parts: the Global Brightness Adjustment (GBA) branch, the Local Feature Extraction (LFE) branch, and the SNR Fusion (SNF) branch.

#### Global brightness adjustment

Zhang et al.<sup>47</sup> have demonstrated that enhancing the V channel of an image in the HSV space can represent the processes of contrast and brightness enhancement while also minimizing noise and color distortion. Additionally, Guo et al.<sup>6</sup> prove that iterative application of the following enhancement curve equation effectively extracts enhancement information.

$$LE_m(x) = LE_{m-1}(x) + \omega_m LE_{m-1}(x)(1 - LE_{m-1}(x)) \quad (4)$$

where,  $m$  represents the number of iterations and controls the curvature.  $LE_m(x)$  denotes the enhanced version of the input image, and  $\omega_m$  is a parameter map of the same size as the image.

Inspired by this, our Global Brightness Adjustment module takes the V channel and its histogram from the low-light image as input. It extracts brightness information from the V channel's histogram and then treats it as a trainable curve parameter  $\omega_{0,1,\dots,n}$ . Using an iterative method, we adjust the V channel features to generate the global brightness feature. As shown in Fig. 5 (a), the main component of this branch is a simple multi-layer perceptron with very few parameters. The adjustment process can be expressed as:

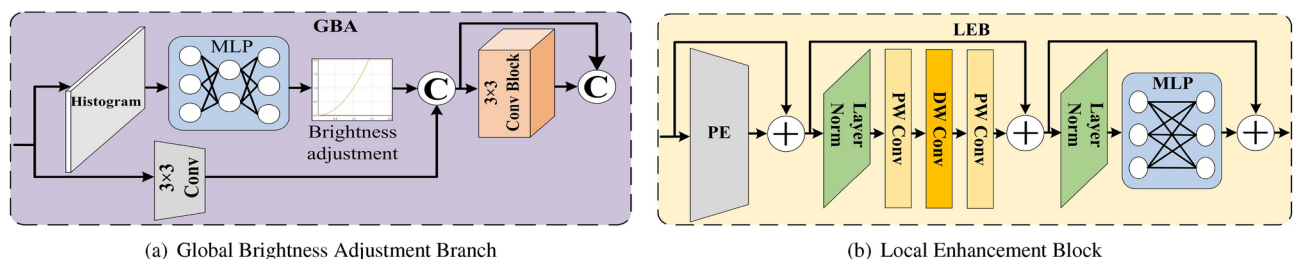
$$\begin{aligned} \omega_{0,1,\dots,n} &= g(h(V)) \\ V_{m+1} &= V_m + \omega_m(V_m - V_m^2) \end{aligned} \quad (5)$$

where,  $g(\cdot)$  represents the multi-layer perceptron part, and  $h(\cdot)$  represents the brightness histogram of the low-light image.

#### Local feature extraction

Transformer is first proposed in the field of natural language processing (NLP)<sup>48</sup>, where its multi-head self-attention mechanism dynamically focuses on different parts of the input sequence in context, enabling outstanding performance in text understanding and generation. Following its success, transformer is gradually introduced to computer vision<sup>49,50</sup>, demonstrating powerful feature extraction capabilities. However, its complex architecture and large parameter scale limit its application in lightweight models.

In the local feature extraction module, our primary goal is to extract features from regions heavily affected by noise. Inspired by the transformer model, we design a transformer-style Local Enhancement Block (LEB). To achieve the lightweight design, we replace the self-attention mechanism with depth-wise convolutions and



**Fig. 5.** Detailed structure of parts in the bilateral enhancement module. **(a)** Global Brightness Adjustment Branch (GBA) mainly consists of an MLP network and high-order curve adjustment. **(b)** Local Enhancement Block (LEB), mainly comprises layer normalization, depth-wise separable convolution, and a simple MLP network.

substitute the transformer's feed-forward network with an MLP composed of two  $1 \times 1$  convolutions to enhance feature representation further.

As shown in Fig. 2, the normalized low-light image  $I_N$  first passes through a  $3 \times 3$  depth-wise convolution to expand the channel dimension, producing the input feature  $F_{in}$ . Subsequently,  $F_{in}$  is processed by the local feature extraction (LFE) branch composed of two stacked LEBs. For the lightweight design of the LEB, as shown in Fig. 5 (b), the LEB uses a depth-wise convolution to encode positional information from  $F_{in}$ , which is then connected with  $F_{in}$  via a residual connection to avoid information loss, resulting in  $F_{em}$ . The enhanced local detail feature  $F_{pdp}$  is then extracted using a depth-wise separable convolution network comprising PWConv-DWConv-PWConv with layer normalization. Finally, we use an MLP with layer normalization to further strengthen the feature representation, producing  $F_{leb}$ .

We apply a skip connection between the output features  $F_{leb,2}$  of the stacked LEB and the input features  $F_{in}$  to retain some fundamental information about the original image. The enhancement process can be expressed as:

$$\begin{aligned} F_{in} &= Conv_{3 \times 3}(I_N) \\ F_{leb,2} &= \sum_{k=1}^2 (PWP(Norm(Conv_{3 \times 3}(F_{in,k}) + F_{in,k})) + MLP(Norm(F_{pdp,k}))) \\ F_l &= F_{leb,2} + F_{in} \end{aligned} \quad (6)$$

where,  $F_{in,k}$  represents the input features of the  $k$ -th LEB block, where  $F_{in,1} = F_{in}$ , and  $F_{pdp,k}$  denotes the enhanced features obtained from the  $k$ -th PWConv-DWConv-PWConv operation.  $F_l$  is the local detail features finally output by local feature extraction LFE.

#### SNR fusion

Estimating noise solely from the input image while simultaneously providing a corresponding clean image to estimate the SNR is challenging and significantly increases the model's complexity. To achieve a lightweight design, we use a non-learning-based method to estimate the SNR of the low-light image. As shown in Fig. 2, given a low-light input image  $I_N$ , we first use a mean filtering method to obtain the denoised image  $I_d$ . We then apply a weighted averaging method to both  $I_N$  and  $I_d$  to get the corresponding grayscale images  $I^g$  and  $I_d^g$ , respectively. By calculating the difference between  $I^g$  and  $I_d^g$ , we obtain the noise image  $N$ . Finally, we apply element-wise division to  $I_d^g$  and  $N$  to get the final SNR map  $S$ . The calculation process is expressed as follows:

$$\begin{aligned} I_d &= blur(I_N) \\ I^g &= gray(I_N), I_d^g = gray(I_d) \\ N &= abs(I^g - I_d^g), S = I_d^g / N \end{aligned} \quad (7)$$

Next, we reshape the obtained SNR map to match the dimensions of the global brightness features and local features, and normalize its values to the range  $[0, 1]$ . Finally, we use the refined SNR map  $S'$  as interpolation weights to dynamically fuse global brightness features  $F_g$  and local detail features  $F_l$ . The fusion process can be expressed as:

$$F_{en} = (1 - S') \times F_g + S' \times F_l \quad (8)$$

### Dual-exposure processing module

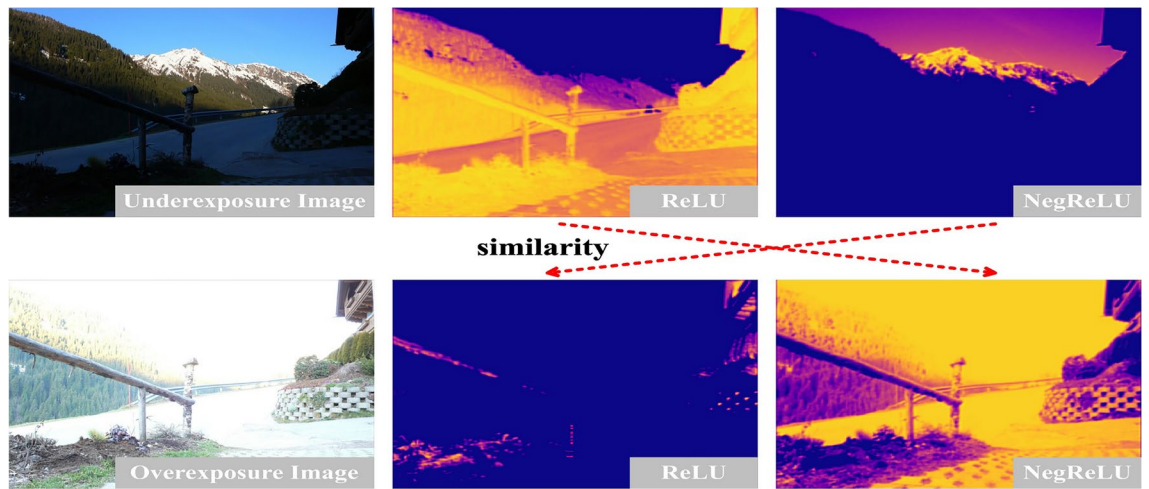
#### Motivation

In the field of image processing, issues of underexposure and overexposure are prevalent and often affect image quality and subsequent computer vision tasks. Traditional image processing techniques typically rely on complex algorithms and parameter adjustments, making it challenging to adaptively handle diverse exposure conditions. With the development of deep learning technology in image processing, new solutions have emerged to address this issue. However, achieving robust handling of different exposure conditions while maintaining a lightweight network remains a challenging task.

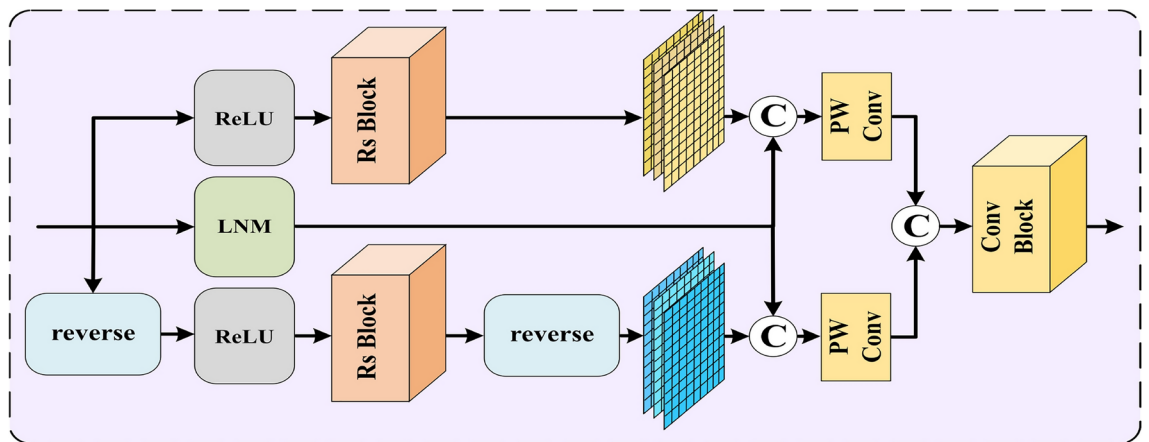
In convolutional neural networks, activation functions play a role in activating certain features, helping the network capture complex characteristics. As shown in Fig. 6, when the network mainly focuses on features of underexposed regions, ReLU and NegReLU functions exhibit differential responses to the two exposure properties, where NegReLU represents the operation of inverting the input values and then applying the ReLU function. Specifically, ReLU tends to process underexposed features, while NegReLU responds more to overexposed features. Additionally, the activation of ReLU for underexposed images and the activation of NegReLU for overexposed images show similarities. Based on this observation, we design the Dual-Exposure Processing (DEP) module, as shown in Fig. 7. It mainly consists of an exposure activation module composed of ReLU and NegReLU activation functions and an exposure learning module with residual networks.

#### Dual-exposure processing

To further improve the robustness of the network under different exposure conditions, we introduce a DEP module after the augmented network. Specifically, for the input feature  $F_{en}$ , we first use ReLU and NegReLU activation functions to obtain features  $F_u$  and  $F_o$  representing underexposed and overexposed properties, respectively. Then, to learn these two features consistently, the exposure learning module processes these



**Fig. 6.** Comparison of activation feature heatmaps for underexposed and overexposed images using ReLU and NegReLU activation functions. When the network activates features of underexposed images, ReLU tends to activate the underexposed parts, whereas NegReLU, in contrast to ReLU, tends to activate the overexposed features. Furthermore, ReLU and NegReLU exhibit similar tendencies in activating features of both underexposed and overexposed images.



**Fig. 7.** Detailed structure of the Dual-Exposure Processing Module. It uses the ReLU activation function to extract features from two exposure properties and then learns these features through residual network blocks. Finally, it employs LNM to obtain exposure-invariant features, which merge with the previously extracted features to produce the final output features.

features through two residual blocks (RsBlock), resulting in  $F'_u$  and  $F'_o$ . Since  $F'_o$  is obtained by inverting ReLU, we need to apply the same inversion process to it before proceeding to the next step to obtain the feature in its original format.

Additionally, to retain more important information from the input feature  $F_{en}$ , we use the LNM module to normalize it and obtain the feature  $F_c$ , which remains invariant to exposure attributes. We then concatenate the three features  $F'_u$ ,  $F'_o$ , and  $F_c$ , so the final feature contains exposure information from both attributes. The whole process is as follows:

$$\begin{aligned} F'_u &= \mathcal{R}(F_u), F'_o = -\mathcal{R}(F_o), F_c = LNM(F_{en}) \\ F''_u &= \mathcal{P}[F'_u, F_c], F''_o = \mathcal{P}[F'_o, F_c], F_{out} = \mathcal{P}[F''_u, F''_o] \end{aligned} \quad (9)$$

where  $\mathcal{R}$  represents the residual network block,  $[-]$  denotes the concatenation operation, and  $\mathcal{P}$  indicates the  $1 \times 1$  convolution layer. Through these operations, the final output features  $F_{out}$  simultaneously contain information from both exposure attributes.



## Loss function

To better enhance the network's performance, this paper employs the loss function that includes not only the commonly used Charbonnier loss function  $L_{char}$  and SSIM structural similarity loss  $L_{ssim}$  but also a color similarity loss function  $L_{color}$ , a luminance similarity loss function  $L_{bright}$ , and a structural similarity loss function  $L_{struct}$ <sup>51</sup>.

### Charbonnier loss function

The  $L_{char}$  is a variant of the  $L_1$  loss function, which, compared to  $L_1$ , includes an additional regularization term  $\epsilon$ . It enhances model performance by approximating the  $L_1$  loss. The gradient near-zero values are prevented from becoming too small due to the presence of  $\epsilon$ , thus avoiding gradient vanishing. It can be expressed as:

$$L_{char} = \sqrt{(\hat{y} - y)^2 + \epsilon^2} \quad (10)$$

where  $\hat{y}$  represents the enhanced image,  $y$  represents the ground truth image, and  $\epsilon$  is a small constant set to  $10^{-3}$  to prevent division by zero.

### Color similarity loss function

The  $L_{color}$  utilizes cosine similarity to measure the hue and saturation differences between two pixels, ensuring that the enhanced image's colors match the reference image more closely. It can be expressed as:

$$L_{color} = 1 - \sum_{i=1, j=1}^{m, n} \cos(E(I_{i,j}), Y_{i,j}) \quad (11)$$

where  $E$  and  $Y$  represent the pixel values of the enhanced image and the reference image, respectively, and the  $\cos(\cdot)$  denotes the cosine similarity between the two vectors. By minimizing the  $L_{color}$ , the network generates enhanced images with hue and saturation closer to those of the ground truth image.

### Brightness similarity loss function

It aims to ensure that the brightness order within each image block of the enhanced image remains consistent with that of the reference image<sup>19</sup>. Specifically, it requires the enhanced image to linearly transform from the noise-free reference image, significantly suppressing noise and improving the quality of the enhanced image. This loss function follows:

$$L_{bright} = 1 - \sum_c \sum_{i=1, j=1}^{r, g, b} \sum_{m, n} \cos(b(E(I_{i,j}^c)) - \min(b(E(I_{i,j}^c)), b(Y(I_{i,j}^c)) - \min(b(Y(I_{i,j}^c)))) \quad (12)$$

where,  $b(\cdot)$  represents image patches centered on  $E$  and  $Y$ . The brightness relationship between them can be expressed as  $b(E) = \lambda b(Y) + \delta$ . Different image patches have different  $\lambda$  and  $\delta$ , and subtracting the minimum value removes the influence of the constant  $\delta$ .

### Structural similarity loss function

It uses gradients to represent structural information, modifying Eq. 12 yields the expression for this loss function:

$$L_{struct} = 1 - \sum_c \sum_{i=1, j=1}^{r, g, b} \sum_{m, n} \cos(b\nabla(E(I_{i,j}^c)) - \min(b\nabla(E(I_{i,j}^c)), b\nabla(Y(I_{i,j}^c)) - \min(b\nabla(Y(I_{i,j}^c)))) \quad (13)$$

To better normalize the brightness distribution, we use the Charbonnier loss function  $L_{low}$  to constrain the normalized low-light image  $I_N$  to be as close as possible to the original low-light image  $I$ . In the experiments, we set the same weight for all loss functions based on empirical observations. Therefore, the total loss is expressed as:

$$L_{total} = L_{char} + L_{ssim} + L_{color} + L_{bright} + L_{struct} + L_{low} \quad (14)$$

## Experiments

### Dataset

In this section, we train our network on the LOL-V1<sup>25</sup> and LOL-V2<sup>26</sup> training datasets and then test it on the corresponding test sets and a mixed dataset. The details of the training and test sets are as follows:

### LOL-V1

It contains 500 image pairs, with 15 used for testing. To highlight the impact of data quantity on the network, we use only 343 image pairs for training. During training, we randomly crop each training image into patches of size  $100 \times 100$  and use data augmentation methods involving random rotation and random flipping to enhance the diversity of the training dataset, effectively reducing and preventing overfitting.

### LOL-V2

This dataset contains 689 image pairs for training and 100 image pairs for testing. During training, we randomly crop each training image into patches of size  $100 \times 100$  and use data augmentation methods involving random rotation and random flipping to enhance the diversity of the training dataset, effectively reducing and preventing overfitting.

### Mixed dataset

This dataset consists of LOL-V1 (15 images)<sup>25</sup>, LIME (10 images)<sup>18</sup>, MF (10 images)<sup>11</sup>, and VV (23 images). It does not have reference images and is used only for evaluating the no-reference metric NIQE<sup>52</sup>.

## Implementation details

The experiment is conducted on a server running Ubuntu 20.04.6 operating system, equipped with an NVIDIA 4090 GPU and a configured PyTorch deep learning framework. During training, the input image size is set to  $100 \times 100$ , with a batch size of 128. To achieve better training results and prevent overfitting, the maximum number of epochs is set to 30,000, and an early stopping mechanism is employed. Through multiple tests, setting the patience value to around 2,000 and the error threshold delta to 0.001 achieves satisfactory training performance. To reduce the number of model parameters, the histogram bin value for the V channel of low-light images is set to 32, which is experimentally validated in the Ablation experiments section. We use the Adam optimization algorithm and find that applying a learning rate adjustment strategy leads to poorer training outcomes. Therefore, both the learning rate and weight decay are set to  $10^{-4}$ . To maintain model accuracy while improving training speed, we set the model saving frequency to every 20 epochs when epoch  $\leq 1,000$ , and every 100 epochs otherwise.

To better compare the low-light enhancement performance of different methods, we use PSNR, SSIM, CIEDE2000<sup>53</sup>, and NIQE<sup>52</sup> as objective evaluation metrics. The PSNR and SSIM measure the peak signal-to-noise ratio and structural similarity of the enhanced images, respectively. Higher values indicate better enhancement effects. The International Commission on Illumination (CIE) introduced CIEDE2000 in 2000 as an improved color difference evaluation metric based on the CIELAB color space. It accounts for the non-uniformity of human visual perception and addresses color perception issues more effectively. Lower values indicate smaller color differences. NIQE is a no-reference image quality assessment metric that measures the perceptual quality of images without needing a reference image. Lower values indicate better perceptual quality of the enhanced images.

## Compared methods

To validate the effectiveness of our method, we compare it with several state-of-the-art low-light enhancement methods, including traditional methods (LIME<sup>18</sup>), unsupervised methods (ZeroDec++<sup>7</sup>, PairLIE<sup>54</sup>), and supervised methods (RetinexNet<sup>25</sup>, MBLEN<sup>8</sup>, KIND<sup>9</sup>, KIND++<sup>55</sup>, IAT<sup>56</sup>, DecNet<sup>57</sup>, FLW-Net<sup>51</sup>).

## Objective evaluation

Tables 1 and 2 present the quantitative comparison of different models on the LOL-V1, LOL-V2, and mixed datasets. We observe that our BiEnNet achieves superior results over other methods in terms of PSNR, SSIM, and CIEDE2000 metrics on both the LOL-V1 and LOL-V2 datasets. Additionally, the number of training samples indeed affects model performance, with a more significant impact on PSNR compared to other metrics. For instance, compared to LOL-V1, training BiEnNet on the LOL-V2 dataset increases the PSNR by nearly 0.6dB (e.g., from 29.06 to 29.66), while the SSIM only increases by about 0.01dB (e.g., from 0.86 to 0.87). The improvements in NIQE and CIEDE2000 are also minimal, decreasing from 3.49 to 3.48 and from 6.16 to 6.10, respectively. Although our method does not achieve the best scores across the board in terms of NIQE, number of parameters, and testing time—for example, on the NIQE metric of the combined dataset, our method yields comparable results to MBLEN, KIND, and KIND++—BiEnNet boasts a relatively low number of parameters (0.1M) and consumes less time during testing. In summary, our BiEnNet attains results that are close to or even better than other methods in overall metrics.

## Visual comparison

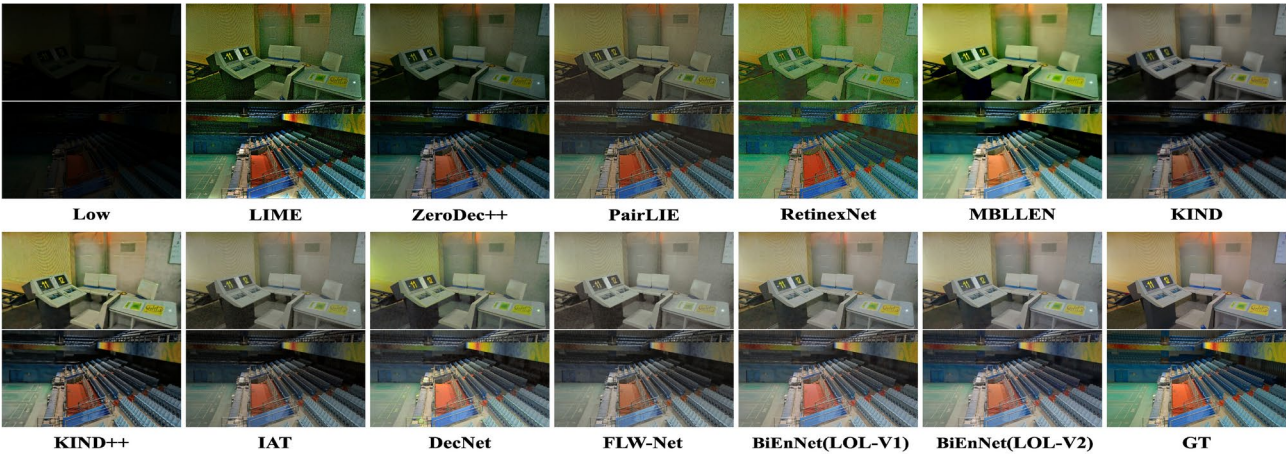
In addition to objective evaluations, we conduct visual comparisons on the LOL-V1, LOL-V2, and mixed datasets to further affirm the effectiveness of our BiEnNet. Fig. 8 and 9 provide the visual comparison results of various methods on the LOL-V1 and LOL-V2 datasets. It is apparent that the low-light images, once enhanced by BiEnNet, exhibit superior improvements in brightness, color, and detail, thus more closely resembling the reference images. As LIME is a traditional enhancement method, it cannot effectively perform targeted enhancements under different conditions, resulting in images with significant noise and darker colors. Moreover, the unsupervised methods ZeroDec++<sup>7</sup> and PairLIE<sup>54</sup>, lacking guidance from reference images during training, do not achieve optimal enhancement effects, especially on the LOL-V1 test dataset. The RetinexNet<sup>25</sup> and MBLEN<sup>8</sup> methods show severe color distortion issues after enhancing low-light images. Fig. 10 presents the visual comparison on the mixed dataset, where our BiEnNet achieves a better balance in detail preservation, color fidelity, and halo artifacts.

Methods	LOL-V1			LOL-V2			Mix Data
	PSNR $\uparrow$	SSIM $\uparrow$	NIQE $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	NIQE $\downarrow$	NIQE $\downarrow$
LIME <sup>18</sup>	17.22	0.50	5.32	15.77	0.46	5.37	4.57
ZeroDec++ <sup>7</sup>	15.35	0.57	7.86	18.49	0.58	8.05	4.53
PairLIE <sup>54</sup>	18.47	0.75	4.25	19.88	0.78	4.34	3.90
RetinexNet <sup>25</sup>	17.86	0.78	6.37	17.37	0.76	9.09	5.68
MBLLEN <sup>8</sup>	17.90	0.70	<b>2.50</b>	18.00	0.72	3.11	<b>3.32</b>
KIND <sup>9</sup>	20.38	0.83	5.45	23.78	0.88	4.96	3.87
KIND++ <sup>55</sup>	21.80	0.84	5.17	22.21	0.84	4.89	3.74
IAT <sup>56</sup>	23.38	0.81	3.92	23.50	0.82	4.19	4.71
DecNet <sup>57</sup>	22.49	0.82	4.51	22.56	0.84	4.83	4.26
FLW-Net <sup>51</sup>	23.84	0.83	4.22	25.71	0.87	4.09	3.93
BiEnNet (LOL-V1)	25.88	0.86	2.75	29.06	0.89	3.08	3.49
BiEnNet (LOL-V2)	<b>26.44</b>	<b>0.87</b>	2.63	<b>29.66</b>	<b>0.90</b>	<b>3.01</b>	3.48

**Table 1.** Quantitative comparison results on the LOL-V1<sup>25</sup>, LOL-V2<sup>26</sup> and mixed datasets. Bold indicates the best results, and italic indicates the second-best results. LIME<sup>18</sup> represents a traditional method, ZeroDec++<sup>7</sup> and PairLIE<sup>54</sup> represent unsupervised learning methods, and the others represent supervised learning methods.

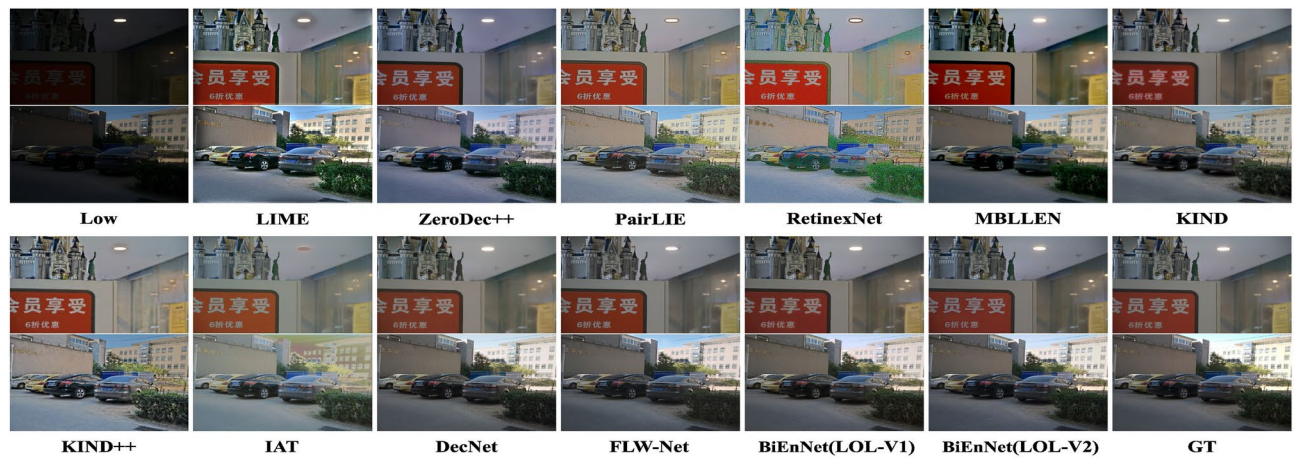
Methods	CIEDE2000 $\downarrow$		Efficiency $\downarrow$	
	LOL-V1	LOL-V2	Params(M)	test time(s)
LIME <sup>18</sup>	14.62	17.63	-	0.19
ZeroDec++ <sup>7</sup>	19.28	14.27	<b>0.01</b>	<b>0.001</b>
PairLIE <sup>54</sup>	11.93	11.23	0.33	0.057
RetinexNet <sup>25</sup>	13.76	18.07	0.40	0.019
MBLLEN <sup>8</sup>	13.72	14.66	20.47	1.981
KIND <sup>9</sup>	9.68	6.74	8.21	0.11
KIND++ <sup>55</sup>	8.51	9.36	8.28	0.12
IAT <sup>56</sup>	7.97	8.17	0.09	0.04
DecNet <sup>57</sup>	8.93	8.87	1.83	0.353
FLW-Net <sup>51</sup>	7.74	6.64	0.02	0.001
BiEnNet (LOL-V1)	6.16	5.03	0.10	0.07
BiEnNet (LOL-V2)	<b>6.10</b>	<b>4.58</b>	0.10	0.07

**Table 2.** The comparison results of color difference metric CIEDE2000<sup>53</sup> and running efficiency on the LOL-V1<sup>25</sup> and LOL-V2<sup>26</sup> datasets. The best and second-best results are marked in Bold and italic, respectively.



**Fig. 8.** Visual comparison of different methods on the LOL-V1<sup>25</sup> dataset, including LIME<sup>18</sup>, ZeroDec++<sup>7</sup>, PairLIE<sup>54</sup>, RetinexNet<sup>25</sup>, MBLLEN<sup>8</sup>, KIND<sup>9</sup>, KIND++<sup>55</sup>, IAT<sup>56</sup>, DecNet<sup>57</sup>, and FLW-Net<sup>51</sup>.





**Fig. 9.** Visual comparison of different methods on the LOL-V2<sup>26</sup> dataset, including LIME<sup>18</sup>, ZeroDec++<sup>7</sup>, PairLIE<sup>54</sup>, RetinexNet<sup>25</sup>, MBLLEN<sup>8</sup>, KIND<sup>9</sup>, KIND++<sup>55</sup>, IAT<sup>56</sup>, DecNet<sup>57</sup>, and FLW-Net<sup>51</sup>.



**Fig. 10.** Visual comparison of different methods on the mixed dataset. The methods include LIME<sup>18</sup>, ZeroDec++<sup>7</sup>, PairLIE<sup>54</sup>, RetinexNet<sup>25</sup>, MBLLEN<sup>8</sup>, KIND<sup>9</sup>, KIND++<sup>55</sup>, IAT<sup>56</sup>, DecNet<sup>57</sup>, and FLW-Net<sup>51</sup>.

## Ablation experiments

### Determination of the bin Value

In this ablation study, we evaluate bin values of 8, 16, 32, 64, and 128. By comparing the changes in model parameters and the PSNR and SSIM performance on the LOL-V1 and LOL-V2 test dataset, we determine the optimal bin value, as shown in Fig. 11.

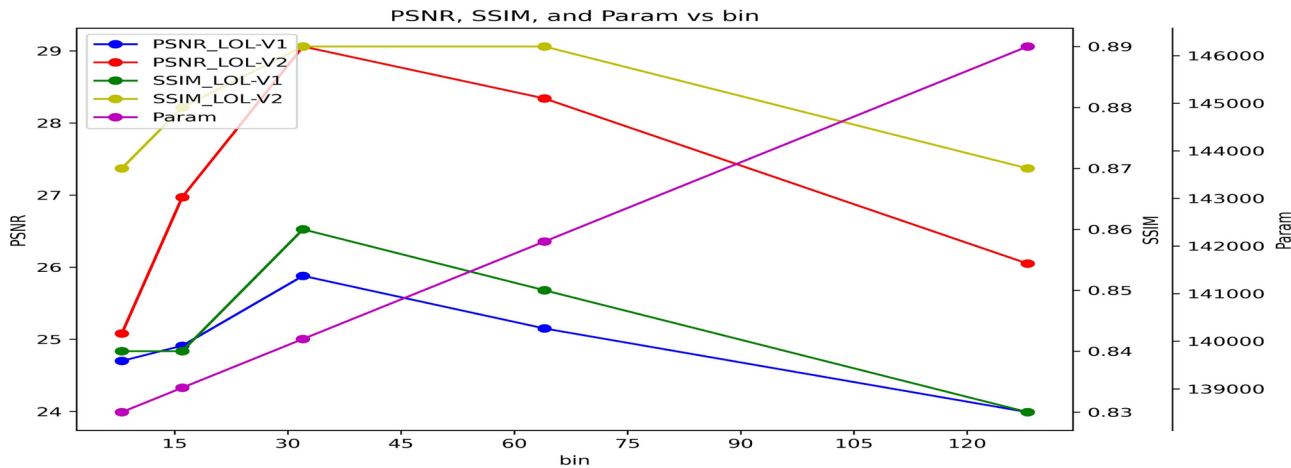
The results indicate that as the bin value increases, both PSNR and SSIM reach their highest values at bin = 32 on both the LOL-V1 and LOL-V2 test datasets. The overall trend shows an initial increase followed by a decline. This phenomenon may be due to the fact that although increasing the number of bins improves model precision, an excessive number of bins can lead to overfitting or parameter redundancy, which decreases accuracy. Moreover, the number of model parameters increases steadily with the bin value. Therefore, to balance model parameters and performance, we choose bin = 32 as the optimal value.

### Demonstration of module effectiveness

We conduct ablation experiments by continuously adding and combining modules in different ways to demonstrate the effectiveness of each module in our proposed method. The entire ablation experiment trains and tests on the LOL-V1 dataset. As shown in Table 3, we use four metrics-PSNR, SSIM, CEIDE2000, and NIQE-in the quantitative comparison experiments.

In Table 3, “1” indicates that the entire network only has a simple Global Brightness Adjustment (GBA) branch. Although it has the lowest scores in all groups, it still achieves good results, demonstrating its effectiveness. “2”, “3”, and “4” add the Luminance Normalization (LNM) module, Dual-Exposure Processing (DEP) module, and the combination of Local Feature Extraction (LFE) and SNR Fusion (SNF) branches to the GBA, respectively. Since LNM and DEP mainly improve the network’s generalization ability, they show similar improvements in PSNR and SSIM metrics. The Bilateral Enhancement Module with SNR Fusion (BSF), composed of LFE, SNF,





**Fig. 11.** Variations in PSNR and SSIM metrics of BiEnNet on the LOL-V1<sup>25</sup> and LOL-V2<sup>26</sup> test datasets under different bin value settings, along with variations in the model’s parameter count.

Index	LNM	DEP	LFE&SNF	PSNR↑	SSIM↑	CIEDE2000↓	NF↓
1	✗	✗	✗	23.84	0.83	7.74	3.89
2	✓	✗	✗	24.13	0.83	7.54	3.61
3	✗	✓	✗	24.16	0.83	7.46	3.58
4	✗	✗	✓	24.85	0.85	7.09	3.22
5	✓	✓	✗	24.74	0.84	7.18	3.39
6	✓	✗	✓	25.24	0.85	6.83	2.95
7	✗	✓	✓	25.35	0.85	6.51	2.83
8	✓	✓	✓	25.88	0.86	6.16	2.75

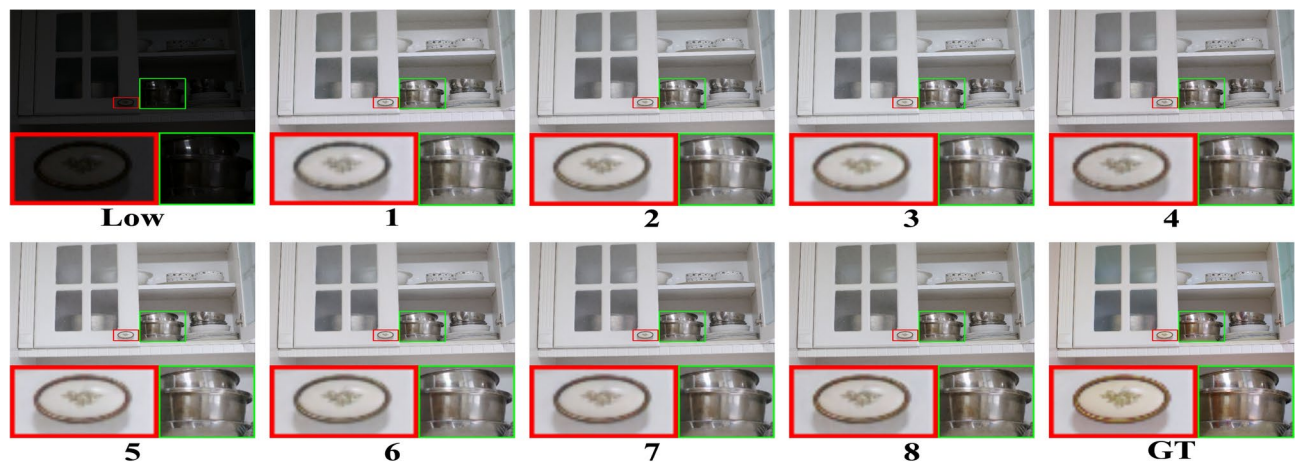
**Table 3.** Visual comparison of different methods on the mixed dataset. The methods include LIME<sup>18</sup>, ZeroDec++<sup>7</sup>, PairLIE<sup>54</sup>, RetinexNet<sup>25</sup>, MBLLEN<sup>8</sup>, KIND<sup>9</sup>, KIND++<sup>55</sup>, IAT<sup>56</sup>, DecNet<sup>57</sup>, and FLW-Net<sup>51</sup>.

and GBA, is the main low-light enhancement part of the entire network. Therefore, “4”, “6”, and “7”, which include LFE and SNF, show significant improvements. The improvement effect of their combinations with LNM and DEP is also similar. Group “8” represents the complete BiEnNet. Due to its improved generalization ability for different degradations and better enhancement capability, this group achieves the best enhancement effect.

As shown in Fig. 12, we also present the visual comparison results of each setting. Group “1” restores overall brightness and contrast, but the enhanced image has unsaturated tones and more noise, especially in the red and green boxed areas. Groups “2” and “3”, which add LNM and DEP, show some improvement in tone restoration but still have significant gap compared to the reference image. Group “4”, “6”, and “7”, which add LFE and SNF, show further improvements in color and detail restoration, especially in the green boxed area. Finally, our complete BiEnNet network, containing all branches, has better enhancement capability. Therefore, “8” achieves an enhancement effect closest to the reference image.

Conclusion

In this paper, we propose a deep learning-based lightweight generalizable low-light enhancement scheme. Specifically, to eliminate the brightness differences in input images, we propose a channel normalization method to obtain a more consistent degradation distribution in the preprocessing stage. In the low-light enhancement stage, we acquire global and local features through a dual-branch enhancement network and effectively fuse the global and local features using the SNR map of low-light images to achieve a better enhancement effect. Finally, to further improve the model’s robustness, we design a dual-exposure processing module in the post-processing state that guides the network to learn features under different exposure conditions simultaneously. Experiments on three public datasets demonstrate the superiority of our method compared to other state-of-the-art methods. However, our method still has certain limitations. For extremely low-light non-synthetic images, most regions contain very little recoverable information, and since our LOL training dataset mainly consists of synthetic low-light images, BiEnNet may produce color distortions that affect overall image quality. In future research, we plan to optimize the network structure while incorporating real low-light image datasets to improve the network’s performance on extremely low-light real-world images. We also aim to explore its potential applications in downstream image processing tasks.



**Fig. 12.** Visual comparison results on the LOL-V1<sup>25</sup> dataset under different settings. The main comparison focuses on the red and green boxed areas. The visual effect when using all branches together is the closest to the reference image.

### Data availability

The data that support the findings of this study are available from the corresponding author, [Z.H.], upon reasonable request.

Received: 4 September 2024; Accepted: 28 November 2024

Published online: 30 November 2024

### References

- Liu, Q., Dong, Y. & Li, X. Multi-stage context refinement network for semantic segmentation. *Neurocomputing* **535**, 53–63. <https://doi.org/10.1016/j.neucom.2023.03.006> (2023).
- Xu, J., Xiong, Z. & Bhattacharyya, S. P. PIDNet: A Real-time Semantic Segmentation Network Inspired by PID Controllers. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 19529–19539, <https://doi.org/10.1109/CVPR52729.2023.01871> (2023).
- Wang, W., Li, S., Shao, J. & Jumahong, H. LKC-Net: Large kernel convolution object detection network. *Scientific Reports* **13**, 9535. <https://doi.org/10.1038/s41598-023-36724-x> (2023).
- Liu, H., Jin, F., Zeng, H., Pu, H. & Fan, B. Image Enhancement Guided Object Detection in Visually Degraded Scenes. *IEEE Transactions on Neural Networks and Learning Systems* 1–14, <https://doi.org/10.1109/TNNLS.2023.3274926> (2023).
- Bhutto, J. A., Khan, A. & Rahman, Z. Image Restoration with Fractional-Order Total Variation Regularization and Group Sparsity. *Mathematics* **11**, 3302. <https://doi.org/10.3390/math11153302> (2023).
- Guo, C. et al. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1777–1786, <https://doi.org/10.1109/CVPR42600.2020.00185> (2020).
- Li, C., Guo, C. & Loy, C. C. Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**, 4225–4238. <https://doi.org/10.1109/TPAMI.2021.3063604> (2022).
- Lv, F., Lu, F., Wu, J. & Lim, C. MBLEN: Low-Light Image/Video Enhancement Using CNNs. In *British Machine Vision Conference* (2018).
- Zhang, Y., Zhang, J. & Guo, X. Kindling the Darkness: A Practical Low-light Image Enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, 1632–1640, <https://doi.org/10.1145/3343031.3350926> (Association for Computing Machinery, New York, NY, USA, 2019).
- Rahman, Z., Pu, Y.-F., Aamir, M. & Wali, S. Structure revealing of low-light images using wavelet transform based on fractional-order denoising and multiscale decomposition. *The Visual Computer* **37**, 865–880. <https://doi.org/10.1007/s00371-020-01838-0> (2021).
- Fu, X. et al. A fusion-based enhancing method for weakly illuminated images. *Signal Processing* **129**, 82–96. <https://doi.org/10.1016/j.sigpro.2016.05.031> (2016).
- Chen, Y., Wen, C., Liu, W. & He, W. A depth iterative illumination estimation network for low-light image enhancement based on retinex theory. *Scientific Reports* **13**, 19709. <https://doi.org/10.1038/s41598-023-46693-w> (2023).
- Rahman, Z., Yi-Fei, P., Aamir, M., Wali, S. & Guan, Y. Efficient Image Enhancement Model for Correcting Uneven Illumination Images. *IEEE Access* **8**, 109038–109053. <https://doi.org/10.1109/ACCESS.2020.3001206> (2020).
- Rahman, Z., Aamir, M., Bhutto, J. A., Hu, Z. & Guan, Y. Innovative Dual-Stage Blind Noise Reduction in Real-World Images Using Multi-Scale Convolutions and Dual Attention Mechanisms. *Symmetry* **15**, 2073. <https://doi.org/10.3390/sym15112073> (2023).
- Pisano, E. D. et al. Contrast Limited Adaptive Histogram Equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital Imaging* **11**, 193. <https://doi.org/10.1007/BF03178082> (1998).
- Zuiderveld, K. Contrast Limited Adaptive Histogram Equalization. In *Graphics Gems*, 474–485, <https://doi.org/10.1016/B978-0-12-336156-1.50061-6> (Elsevier, 1994).
- Lee, C., Lee, C. & Kim, C.-S. Contrast Enhancement Based on Layered Difference Representation of 2D Histograms. *IEEE Transactions on Image Processing* **22**, 5372–5384. <https://doi.org/10.1109/TIP.2013.2284059> (2013).
- Guo, X., Li, Y. & Ling, H. LIME: Low-Light Image Enhancement via Illumination Map Estimation. *IEEE Transactions on Image Processing* **26**, 982–993. <https://doi.org/10.1109/TIP.2016.2639450> (2017).
- Wang, S., Zheng, J., Hu, H.-M. & Li, B. Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. *IEEE Transactions on Image Processing* **22**, 3538–3548. <https://doi.org/10.1109/TIP.2013.2261309> (2013).

20. Fu, X., Zeng, D., Huang, Y., Zhang, X.-P. & Ding, X. A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2782–2790, <https://doi.org/10.1109/CVPR.2016.304> (2016).
21. Li, M., Liu, J., Yang, W., Sun, X. & Guo, Z. Structure-Revealing Low-Light Image Enhancement Via Robust Retinex Model. *IEEE Transactions on Image Processing* **27**, 2828–2841. <https://doi.org/10.1109/TIP.2018.2810539> (2018).
22. Rahman, Z., Bhutto, J. A., Aamir, M., Dayo, Z. A. & Guan, Y. Exploring a radically new exponential Retinex model for multi-task environments. *Journal of King Saud University - Computer and Information Sciences* **35**, 101635. <https://doi.org/10.1016/j.jksuci.2023.101635> (2023).
23. Xu, X., Wang, R., Fu, C.-W. & Jia, J. SNR-Aware Low-light Image Enhancement. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17693–17703, <https://doi.org/10.1109/CVPR52688.2022.01719> (2022).
24. Wu, Y. et al. Learning Semantic-Aware Knowledge Guidance for Low-Light Image Enhancement. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1662–1671, <https://doi.org/10.1109/CVPR52729.2023.00166> (2023).
25. Wei, C., Wang, W., Yang, W. & Liu, J. Deep Retinex Decomposition for Low-Light Enhancement, <https://doi.org/10.48550/arXiv.1808.04560> (2018). arXiv: 1808.04560.
26. Yang, W., Wang, W., Huang, H., Wang, S. & Liu, J. Sparse Gradient Regularized Deep Retinex Network for Robust Low-Light Image Enhancement. *IEEE Transactions on Image Processing* **30**, 2072–2086. <https://doi.org/10.1109/TIP.2021.3050850> (2021).
27. Jiang, Y. et al. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. *IEEE Transactions on Image Processing* **30**, 2340–2349. <https://doi.org/10.1109/TIP.2021.3051462> (2021).
28. Wu, W. et al. URetinex-Net: Retinex-based Deep Unfolding Network for Low-light Image Enhancement. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5891–5900, <https://doi.org/10.1109/CVPR52688.2022.00581> (2022).
29. Ioffe, S. & Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, <https://doi.org/10.48550/arXiv.1502.03167> (2015). arXiv: 1502.03167.
30. Wu, Y. & He, K. Group Normalization. *International Journal of Computer Vision* **128**, 742–755. <https://doi.org/10.1007/s11263-019-01198-w> (2020).
31. Schaffer, C. Selecting a classification method by cross-validation. *Machine Learning* **13**, 135–143. <https://doi.org/10.1007/BF00993106> (1993).
32. Zhou, K., Liu, Z., Qiao, Y., Xiang, T. & Loy, C. C. Domain Generalization: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**, 4396–4415. <https://doi.org/10.1109/TPAMI.2022.3195549> (2023).
33. Zhang, Y. et al. Underwater Image Enhancement Using Deep Transfer Learning Based on a Color Restoration Model. *IEEE Journal of Oceanic Engineering* **48**, 489–514. <https://doi.org/10.1109/OJE.2022.3227393> (2023).
34. Park, S., Yoo, J., Cho, D., Kim, J. & Kim, T. H. Fast Adaptation to Super-Resolution Networks via Meta-learning. In Vedaldi, A., Bischof, H., Brox, T. & Frahm, J.-M. (eds.) *Computer Vision – ECCV 2020*, 754–769, [https://doi.org/10.1007/978-3-030-58583-9\\_45](https://doi.org/10.1007/978-3-030-58583-9_45) (Springer International Publishing, Cham, 2020).
35. Zheng, S. & Gupta, G. Semantic-Guided Zero-Shot Learning for Low-Light Image/Video Enhancement. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 581–590, <https://doi.org/10.1109/WACVW54805.2022.00064> (2022).
36. Ma, L., Ma, T., Liu, R., Fan, X. & Luo, Z. Toward Fast, Flexible, and Robust Low-Light Image Enhancement. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5627–5636, <https://doi.org/10.1109/CVPR52688.2022.00555> (2022).
37. Yang, S., Ding, M., Wu, Y., Li, Z. & Zhang, J. Implicit Neural Representation for Cooperative Low-light Image Enhancement. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 12872–12881, <https://doi.org/10.1109/ICCV51070.2023.01187> (2023).
38. Fan, X. et al. Adversarially Adaptive Normalization for Single Domain Generalization. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8204–8213, <https://doi.org/10.1109/CVPR46437.2021.00811> (2021).
39. Reza, A. M. Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement. *Journal of VLSI signal processing systems for signal, image and video technology* **38**, 35–44. <https://doi.org/10.1023/B:VLSI.0000028532.53893.82> (2004).
40. Li, F. et al. Gamma-enhanced Spatial Attention Network for Efficient High Dynamic Range Imaging. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1031–1039, <https://doi.org/10.1109/CVPRW56347.2022.00116> (2022).
41. Mertens, T., Kautz, J. & Van Reeth, F. Exposure Fusion: A Simple and Practical Alternative to High Dynamic Range Photography. *Computer Graphics Forum* **28**, 161–171. <https://doi.org/10.1111/j.1467-8659.2008.01171.x> (2009).
42. Zhang, Q., Nie, Y., Zhang, L. & Xiao, C. Underexposed Video Enhancement via Perception-Driven Progressive Fusion. *IEEE Transactions on Visualization and Computer Graphics* **22**, 1773–1785. <https://doi.org/10.1109/TVCG.2015.2461157> (2016).
43. Afifi, M., Derpanis, K. G., Ommer, B. & Brown, M. S. Learning Multi-Scale Photo Exposure Correction. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9153–9163, <https://doi.org/10.1109/CVPR46437.2021.00904> (2021).
44. Rahman, Z. et al. Diverse image enhancer for complex underexposed image. *Journal of Electronic Imaging* **31**, 041213. <https://doi.org/10.1117/1.JEL.31.4.041213> (2022).
45. Rahman, Z. et al. Efficient Contrast Adjustment and Fusion Method for Underexposed Images in Industrial Cyber-Physical Systems. *IEEE Systems Journal* **17**, 5085–5096. <https://doi.org/10.1109/JSYST.2023.3262593> (2023).
46. Ulyanov, D., Vedaldi, A. & Lempitsky, V. Instance Normalization: The Missing Ingredient for Fast Stylization, <https://doi.org/10.48550/arXiv.1607.08022> (2017). arXiv: 1607.08022.
47. Zhang, Y., Di, X., Zhang, B., Ji, R. & Wang, C. Better Than Reference in Low-Light Image Enhancement: Conditional Re-Enhancement Network. *IEEE Transactions on Image Processing* **31**, 759–772. <https://doi.org/10.1109/TIP.2021.3135473> (2022).
48. Vaswani, A. et al. Attention Is All You Need, <https://doi.org/10.48550/arXiv.1706.03762> (2023). arXiv: 1706.03762.
49. Dosovitskiy, A. et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, <https://doi.org/10.48550/arXiv.2010.11929> (2021). arXiv: 2010.11929.
50. Yang, S., Zhou, D., Cao, J. & Guo, Y. Rethinking Low-Light Enhancement via Transformer-GAN. *IEEE Signal Processing Letters* **29**, 1082–1086. <https://doi.org/10.1109/LSP.2022.3167331> (2022).
51. Zhang, Y. et al. Simplifying Low-Light Image Enhancement Networks with Relative Loss Functions, <https://doi.org/10.48550/arXiv.2304.02978> (2023). arXiv: 2304.02978.
52. Mittal, A., Soundararajan, R. & Bovik, A. C. Making a “Completely Blind” Image Quality Analyzer. *IEEE Signal Processing Letters* **20**, 209–212. <https://doi.org/10.1109/LSP.2012.2227726> (2013).
53. Luo, M. R., Cui, G. & Rigg, B. The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application* **26**, 340–350. <https://doi.org/10.1002/col.1049> (2001).
54. Fu, Z. et al. Learning a Simple Low-Light Image Enhancer from Paired Low-Light Instances. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22252–22261, <https://doi.org/10.1109/CVPR52729.2023.02131> (2023).
55. Zhang, Y., Guo, X., Ma, J., Liu, W. & Zhang, J. Beyond Brightening Low-light Images. *International Journal of Computer Vision* **129**, 1013–1037. <https://doi.org/10.1007/s11263-020-01407-x> (2021).

56. Cui, Z. *et al.* You Only Need 90K Parameters to Adapt Light: A Light Weight Transformer for Image Enhancement and Exposure Correction, <https://doi.org/10.48550/arXiv.2205.14871> (2022). [arXiv: 2205.14871](https://arxiv.org/abs/2205.14871).
57. Liu, X., Xie, Q., Zhao, Q., Wang, H. & Meng, D. Low-Light Image Enhancement by Retinex-Based Algorithm Unrolling and Adjustment. *IEEE Transactions on Neural Networks and Learning Systems* 1–14, <https://doi.org/10.1109/TNNLS.2023.3289626> (2023).

## Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 62472145 and 62273292), the Natural Science Foundation of Henan Province (No. 242300420284) and the Fundamental Research Funds for the Universities of Henan Province (No. NSFRF240820).

## Author contributions

J.W. and Z.H. conceived the experiment(s), J.W., S.H. and Z.H. conducted the experiment(s), S.H., S.Z. and Y.Q. analysed the results. All authors reviewed the manuscript.

## Declarations

## Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The corresponding author states that there is no conflict of financial or non-financial interests. We would like to declare that the work described was original research that has not been published previously. It is not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

## Additional information

**Correspondence** and requests for materials should be addressed to Z.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024