



IATN: illumination-aware two-stage network for low-light image enhancement

Shuying Huang¹ · Huiying Dong² · Yong Yang² · Yingzhi Wei² · Mingyang Ren² · Shuzhao Wang²

Received: 2 December 2023 / Revised: 30 December 2023 / Accepted: 14 January 2024 / Published online: 17 February 2024
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

Abstract

Images captured in low brightness environment have issues with low contrast and noise due to uneven lighting, which can seriously affect the accuracy of high-level computer vision tasks. Currently, most enhancement methods still suffer from color distortion and noise amplification. To overcome these issues, this paper proposes an illumination-aware two-stage network (IATN) for low-light image enhancement. In the first stage, a tiny illumination estimation network based on Retinex theory is constructed to generate a coarse enhanced image. In the second stage, an illumination-aware correction network (IACN) is designed by building an illumination map to guide the reconstruction of features, which can reduce color distortion and suppress noise in the results obtained in the first stage, thereby obtaining refined enhancement results. In IACN, considering the exposure difference in different regions of the image caused by uneven lighting, multiple illumination-aware modules are constructed to correct features at different scales by utilizing the long-range dependence of features. Numerous experiments conducted on public benchmark datasets have shown that our IATN generates enhanced images that are more natural, colorful, and superior to some state-of-the-art methods. The source code of this work will be available on GitHub.

Keywords Low-light image enhancement · Retinex theory · Two-stage network · Illumination aware · Multi-scale feature enhancement

1 Introduction

Low-light image enhancement (LLIE) methods aim to restore the brightness of images captured in low-light environments and suppress noise caused by uneven lighting. Low-light images are rather common phenomena in real-world scenarios, and insufficient lighting leads to issues such as detail loss and low contrast in captured images, which greatly reduces the visual quality of the image. These degradations will not only have negative effect on human visual perception, but also will be not conducive to the subsequent computer vision tasks designed for high-quality images, such as face recognition [1], object detection [2] and semantic segmentation [3]. In practical applications, we can improve the brightness of images through digital image processing technology, as shown in Fig. 1.

In recent years, a large number of studies on LLIE have been conducted. Traditional LLIE methods can be roughly categorized into two types: histogram equalization (HE)-based methods [4] and Retinex model-based methods [5, 6]. Although these methods can improve the brightness of the image, the enhanced images obtained usually have serious color deviation and edge blur issues.

Recently, due to the great success of deep learning technology in computer vision tasks, many deep learning-based LLIE methods have proposed, which are mainly divided into two categories: end-to-end frameworks [7–10] and Retinex-based frameworks [11–13]. The end-to-end frameworks are to learn the direct mapping function from a low-light image to the corresponding normal-light image. However, the results of such methods often suffer from color distortion and noise, and the constructed models lack interpretability, making it difficult to adapt to images in real low brightness scenes. Retinex-based frameworks typically decompose a low-light image into illumination and reflection maps and generate a normal-light images by adjusting the illumination and reflection components separately. However, the training process of such frameworks is cumbersome and time-consuming, and

✉ Yong Yang
greatyangy@126.com

¹ School of Software, Tiangong University, Tianjin, China

² School of Computer Science and Technology, Tiangong University, Tianjin, China



Fig. 1 Visualization comparison of low-light image and enhancement images

the quality of enhancement images largely depends on the accuracy of low-light image decomposition.

To tackle the above issues, we propose an illumination-aware two-stage network (IATN) based on Retinex theory for LLIE, which gradually achieves image enhancement by constructing two stages: illumination estimation and feature correction. In the first stage, to reduce the dependence on image decomposition and the computational complexity of the model, we construct a tiny network based on Retinex theory, i.e., illumination estimation network (IEN), to generate a coarse enhancement result. In the second stage, to address the issues of noise and color distortion in the result from the first stage, an illumination-aware correction network (IACN) consisting of an illumination feature extraction branch and a feature correction branch is constructed to obtain a refine enhancement result. In IACN, considering the diversity of lighting conditions, an initial illumination map is first directly obtained from the input low-light image and is sent to the illumination feature extraction branch to extract illumination features at different scales. Then, the U-shape feature correction branch containing multiply illumination-aware modules (IAMs) is designed to achieve correction and restoration of features at different scales. IAM is the core module for feature restoration, which guides the network in achieving noise removal and color correction by establishing the autocorrelation of image features in the channel dimension and the correlation between image features and illumination guidance features in the spatial dimension.

Overall, our contributions are summarized as follows:

- (1) An IATN for low-light image enhancement is proposed by constructing an IEN and an IACN, in order to gradually obtain the enhanced results with better visual effects.
- (2) In the first stage, a tiny IEN based on Retinex theory is constructed by learning an illumination map to obtain a coarse enhancement result.
- (3) In the second stage, an IACN is constructed to obtain a refined enhancement result by constructing an illumination feature extraction branch and a feature correction branch. In the feature correction branch, an IAM is designed to achieve restoration of features by

establishing the correction between the illumination features and image features.

- (4) Extensive experiments on public low-light image datasets show that the proposed network achieves better performance both subjectively and objectively compared to some state-of-the-art methods.

2 Related works

2.1 Traditional LLIE methods

The traditional LLIE methods are roughly divided into two categories: histogram equalization (HE)-based methods and Retinex model-based methods.

Histogram-based methods enhance the image by extending the dynamic range of pixel values. For example, Celik et al. [14] enhanced the contrast of input image by mapping the diagonal elements of the initial histogram to the diagonal elements of the target histogram. Lee et al. [15] applied the layered difference representation of 2D histograms to amplify the gray-level differences between adjacent pixels. Although such methods can enhance the brightness of an image, it often fails to achieve the enhancement effect due to its simple use of mathematical methods without considering the formation model of the image.

The Retinex-based models follow a common assumption according to the Retinex theory, i.e., an image can be represented by the product of an illumination layer describing the illumination distribution of the scene and a reflectance layer describing the surface property of the object. The Retinex-based method obtains enhancement results by separately processing the decomposed reflectance and illumination components. For example, Fu et al. [16] proposed a weighted variational model to simultaneously estimate the illumination and the reflectance maps from an observed image. Hao et al. [17] proposed a novel Retinex-based LLIE method performed in a semi-decoupled way. However, this enhancement problem is a “pathological” problem, and due to imprecise prior assumptions, there is often overexposure in the enhancement results.

2.2 Learning-based LLIE methods

Due to the powerful learning ability of deep networks, the learning-based LLIE method has gradually become the mainstream method for LLIE. For example, Wei et al. [11] proposed a two-stage Retinex-based method called RetinexNet. The first stage decomposes the input image into reflectance and illumination maps, and the second stage obtains the enhanced result by adjusting the illumination map. Inspired by RetinexNet, Zhang et al. proposed two refined methods, called KinD [18] and KinD++ [13], which includes three sub-networks: decomposition-net, restoration-net and adjustment-net. Guo et al. [19] proposed a LLIE framework based on the divide-and-rule principle, which converts the image from RGB space into a luminance-chrominance space and designs an adjustable noise suppression network to eliminate noise in the brightened luminance. These methods have the problem of complex and time-consuming training, so some researchers have further studied end-to-end methods. For example, Liang et al. [20] designed a DCP-guided hierarchical dynamic mechanism for end-to-end LLIE. Yang et al. [21] presented a U-shaped encoder–decoder network based on multi-scale feature complementation. Fu et al. [22] designed an unsupervised network based on Retinex theory for learning adaptive priors from low-light image pairs, which can generate clean images through consistent constraints on reflectance. However, due to the inability to establish long-range dependencies of features in CNN-based methods, the issues of noise and color distortion have not been effectively addressed.

Due to its superior performance in capturing long-range features, Transformer is widely used in computer vision tasks. The Transformer [23] was first proposed for the field of natural language processing, and its superior performance has motivated many researchers to introduce it to vision tasks [24–26]. In image restoration, Zhang et al. [27] proposed a Structure and Texture-Aware Network, in which the structure sub-network is composed of stacked Transformer module, while the texture sub-network is composed of stacked central difference convolution modules. Xu et al. [9] proposed an SNR-aware CNN-Transformer hybrid network, called SNR-Net, which enhances areas with low signal-to-noise ratio using Transformer and enhances areas with high signal-to-noise ratio using convolution operations. However, these methods are accompanied by a huge computational burden and are difficult to apply high-resolution images.

3 Proposed method

3.1 Motivation and overview

Based on the Retinex theory, an observation image can be decomposed into two components: a reflection map and an illumination map. The reflection map represents the reflection component of the inherent properties of the object itself, which is not affected by light. The illuminance map reflects the information of light intensity and also determines the dynamic range size of the grayscale values of all pixels in the observation image. Under low illumination, due to uneven lighting, there may also be noise in the observation image. Therefore, a low-light image is affected by the light intensity and noise, and its degradation model can be defined as follows:

$$I_{low} = R \odot L + N \quad (1)$$

where I_{low} stands for the low-light image, R stands for reflection map, L is the illumination map, N represents the noise, and \odot represents the element-wise multiplication.

According to formula (1), an IATN for low-light image enhancement is proposed, as shown in Fig. 2. The first stage network is constructed based on the first term to the right of the equal sign in formula (1) and obtain a coarse enhancement result by estimating the illumination map \bar{L} . Due to the uneven illumination intensity, there are often issues of noise amplification and color distortion in the result \bar{R} from the first stage. Therefore, an IACN in the second stage network is constructed to obtain a refined enhancement result \bar{I} .

Below, we will provide a detailed introduction to the construction of the two stages in the proposed IATN.

3.2 Illumination estimation network (IEN)

In the first stage, based on the Retinex theory, an IEN is constructed to learn an illumination map from the input low-light image and obtain a coarse enhancement result by utilizing the inverse formula $\bar{R} = I / \bar{L}$. Because the illumination map reflects the light intensity information in the environment and does not contain specific content, there is no high-frequency component in an ideal illumination map. Therefore, instead of learning complex features, the IEN requires a very small architecture to learn an illumination map, as shown in Fig. 2. In the IEN, three simple convolutional layers are adopted,

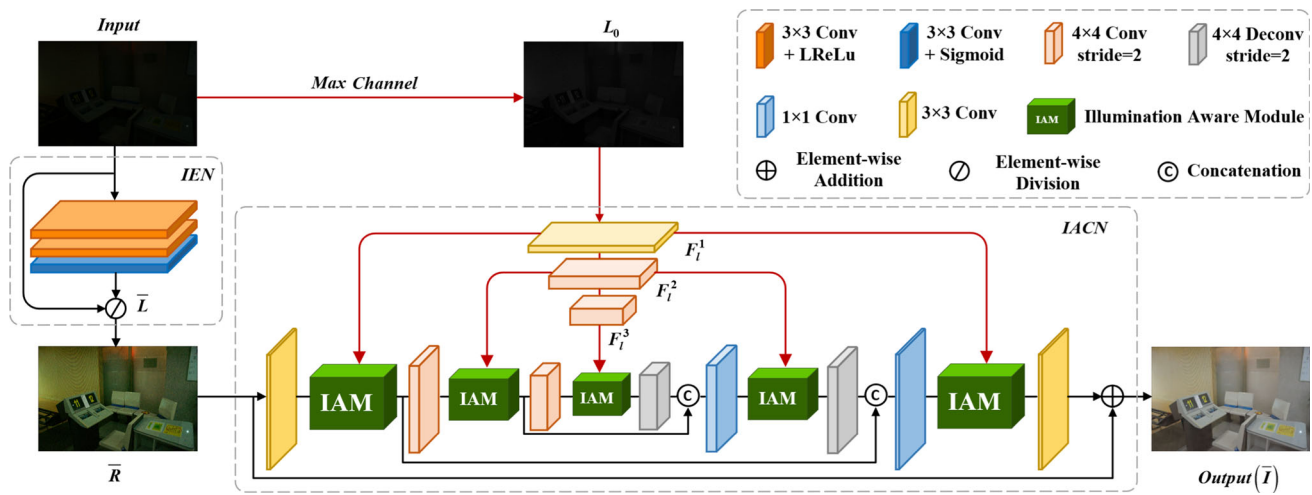


Fig. 2 The overall framework of the proposed IATN

each of which contains a convolutional operation and an activation function to extract features and increase the non-linearity of the network. A sigmoid function in the final convolutional layer is used to compress the values in the illumination map to the range of [0,1].

To make the learned illumination map more accurate, a dual-constraint loss function is defined, which includes two loss terms to constrain the generated illumination map and coarse enhancement image, respectively. The first loss term is defined as the structure-aware TV loss [28] and can be expressed as:

$$L_{is} = \|\nabla \bar{L} \odot \exp(-\lambda_g \nabla I_{GT})\| \tag{2}$$

where ∇ denotes the gradient operator, I_{GT} denotes the ground-truth (GT) image, λ_g denotes the coefficient balancing the strength of structure awareness, and $\exp(\cdot)$ denotes an exponential function. The second loss term is defined as a mean squared error (MSE) loss, which is used to evaluate the difference between the coarse enhancement image and the corresponding GT image. It can be expressed as:

$$L_{mse} = \|\bar{R} - I_{GT}\|_2^2 \tag{3}$$

Based on Eqs. 2 and 3, the dual-constraint loss function of IEN can be expressed as:

$$L_{ie} = L_{is} + L_{mse} \tag{4}$$

3.3 Illumination-aware correction network (IACN)

Although the coarse enhancement result already has good visual effects, it is still affected by noise amplification and color distortion due to uneven lighting. The degree of these

two degradation factors is affected by uneven illumination, and noise and color distortion become more pronounced in darker areas. Therefore, in the second stage shown in Fig. 2, an IACN consisting of two branches, namely the illumination feature extraction branch and a feature correction branch, is designed to remove noise and correct color in the coarse enhancement result \bar{R} , in order to obtain the refined enhancement image with better visual effect.

In the illumination feature extraction branch, the illumination features at different scales are generated and guide the restoration of features in the feature correction branch. The specific operation is as follows. Firstly, considering the illuminance differences of input low-light images, the initial illuminance map used to extract illumination features is directly obtained from the maximum values of the three channels in the input low-light image. Then, three convolutional layers containing a 3×3 convolution and two 4×4 convolutions with a stride of 2 are used to obtain illumination feature maps $F_i^i (i = 1, 2, 3)$ at different scales, where i represents the scales. The number of channels at each scale in the illumination feature extraction branch corresponds to the number of channels in the feature correction branch at the same scale. Finally, these feature maps are sent to the corresponding layers of the feature correction branch.

In the feature correction branch, a U-shaped network containing three feature correction layers is designed, which corrects features of different scales through the guidance of illumination features. In each feature correction layer, multiple IAM are used to achieve feature enhancement by establishing the correlation between illuminance guidance features and features extracted from \bar{R} . The 4×4 convolution or deconvolution with a stride of 2 is used to downsampling or upsampling the feature maps while expanding or compressing the channel dimension. The structure of IAM shown in Fig. 3 will be described in detail below.

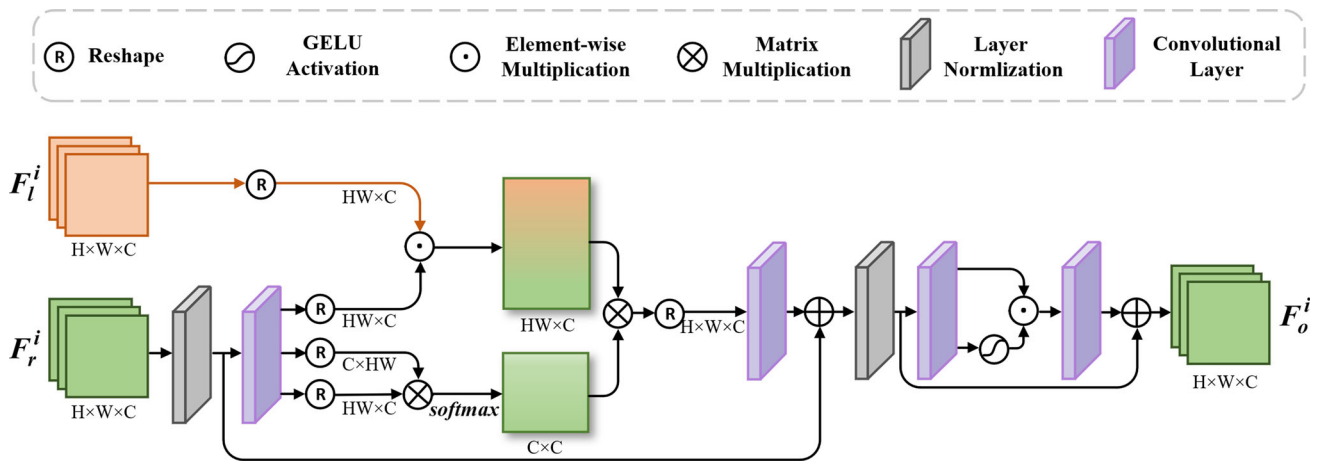


Fig. 3 The structure of IAM

IAM receives the feature maps F_r^i from \bar{R} and guidance feature maps F_l^i to achieve the enhancement of image features by constructing correlations between the two features. The execution process of IAM is as follows. First, the feature maps F_r^i undergo a layer normalization and convolutional layer to generate query (Q), key (K), value (V) projections. This process can be defined as:

$$\begin{aligned}
 W &= WQ(LN(F_r^i)), K = WK(LN(F_r^i)), \\
 V &= WV(LN(F_r^i))
 \end{aligned}
 \tag{5}$$

where WQ, WK, WV represent the convolution layers, and LN is the layer normalization.

Then, to reduce the computational complexity of the network and capture global contextual relationships between pixels, Q and K projections are reshaped to generate a transposed-attention map across feature dimensions through dot-product interaction. In addition, to perceive the impact of illumination on different regions, the guidance feature maps are reshaped and multiplied pixel by pixel with the reshaped V projection to generate an illumination-aware attention map on the spatial dimension. The specific operations are defined as follows:

$$\begin{aligned}
 A &= \text{softmax}(\hat{Q} \times \hat{K} / \alpha) \\
 IA &= \hat{F}_l^i \odot \hat{V}
 \end{aligned}
 \tag{6}$$

where $A \in R^{C \times C}$ represents the transposed-attention map, $\hat{Q}, \hat{K}, \hat{V}$ represents reshaped feature maps from Q, K, V separately, α represents a learnable parameter that adaptively scales the matrix multiplication, \hat{F}_l^i represents the reshaped illumination feature maps from F_l^i and $IA \in R^{H \times W \times C}$ represents the illumination-aware attention map.

Next, the two attention maps are multiplied to achieve interaction of local and nonlocal features, and the generated

feature maps are aggregated with the input feature map F_r^i through a residual operation. Finally, the generated feature maps undergo a gated-convolutional feed-forward network [29] to obtain the output of the IAM, which is the enhanced features. This process is defined as follows:

$$F_o^i = FN(A \times IA + F_r^i)
 \tag{7}$$

where FN denotes gated-convolutional feed-forward network, and F_o^i is the output of IAM.

For the training of IACN, we define a loss function that includes two loss terms: the Charbonnier loss [30] and structural similarity loss [31], as shown below:

$$L_{iac} = \lambda \sqrt{\|\bar{I} - I_{GT}\|_F^2 + \epsilon^2} + (1 - \lambda) L_{SSIM}(\bar{I}, I_{GT})
 \tag{8}$$

where L_{SSIM} represents the structural similarity loss, the weighting parameter λ is set to 0.8, and the constant ϵ is set to 0.001.

4 Experiments and analysis

4.1 Implementation details

We implemented our model using PyTorch on a single NVIDIA RTX 3080 GPU. In the first stage of IATN, the batch size is set to 32, and the number of epochs is 10. In the second stage of IATN, the number of IAMs in the IACN is 1, 2, 2, 2 and 2 from the first to the fifth layer, respectively. The batch size is set to 8, and the number of epochs is 2000. In both stages, we augment the data using rotation and horizontal flipping and optimize the networks by the ADAM optimizer. The initial learning rate is set to 0.0001 in both stages, and the learning rate steadily decreases to 0.000001

Table 1 Quantitative results of all comparison methods on LOL and LOL-v2 datasets

Methods	LOL			LOL-v2		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
RetinexNet [11]	16.77	0.425	0.474	16.10	0.407	0.543
GLAD [34]	19.25	0.682	0.321	19.82	0.685	0.315
Zero-DCE [35]	14.86	0.589	0.335	18.06	0.580	0.313
EnlightenGAN [37]	17.48	0.677	0.322	18.64	0.677	0.309
RUAS [36]	18.23	0.717	0.354	15.33	0.493	0.310
URetinex [39]	21.33	<u>0.834</u>	0.121	21.22	<u>0.859</u>	<u>0.099</u>
Night-enhancement [38]	<u>21.52</u>	0.763	0.241	<u>25.51</u>	0.783	0.253
PairLIE [22]	18.47	0.743	0.243	19.88	0.733	0.234
Bread [19]	20.62	0.831	0.164	23.69	0.869	0.156
Ours	23.41	0.846	<u>0.129</u>	28.81	0.901	0.095

The best and second-best results are bolded and underlined, respectively

by the cosine annealing scheme [32] during the training of IACN.

4.2 Dataset and evaluation metrics

We train our model on the LOL dataset [11], which contains 485 pairs of low-/normal-light images for training and 15 pairs for testing. To verify the generalization of the model, we test the trained model on another dataset LOL-v2. The LOL-v2 dataset is the real part of LOL-v2 [12], which is larger and more diverse than LOL, including 689 pairs of low-/normal-light images for training and 100 pairs for testing.

To evaluate the performance of different LLIE methods, we adopt three well-known objective evaluation metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) [31] and Learned Perceptual Image Patch Similarity (LPIPS) [33]. The higher the value of PSNR and SSIM, the better the quality of the image. On the contrary, the lower the value of LPIPS, the better the quality of the image.

4.3 Quantitative and qualitative results

To verify the effectiveness of the proposed method, we compare it with some state-of-the-art (SOTA) methods for low-light enhancement, including RetinexNet [11], GLAD [34], Zero-DCE [35], RUAS [36], EnlightenGAN [37], Night-enhancement [38], URetinex [39], PairLIE [22] and Bread [19]. Note that the results of all comparison methods are reproduced by using the official codes with recommended parameters.

Table 1 shows the quantitative results of all compared methods on the LOL and LOL-v2 datasets, respectively. It is obvious that our method achieves better results on both datasets. On the LOL dataset, the PSNR and SSIM values of

the propose method are higher than those of other comparison methods, while the LPIPS value is only slightly worse than those of URetinex. On the LOL-v2 dataset, our method obtained the best values for all three indicators compared to other methods.

The qualitative results on LOL and LOL-v2 datasets are shown in Figs. 4 and 5, respectively. For ease of observation, we selected and enlarged a small area from the images. As can be seen from the figures, the results of RetinexNet have serious color distortion and noise issues. Zero-DCE, EnlightenGAN and RUAS are unable to effectively restore the brightness of darker areas. The results obtained by Night-enhancement, URetinex and Bread suffer from color distortions compared to the GT. In addition, there is also severe noise in the GLAD results, and the results of Night-enhancement and PairLIE show varying degrees of edge blur. Compared with other methods, our method achieved results with higher visual quality, which are closest to GT.

4.4 Complexity analysis

In this section, we compared the parameters and floating-point operations per second (FLOPs) of all methods. The comparison results are shown in Table 2, from which we can find that the RUAS has the fewest parameters and FLOPs, while the Night-enhancement has the largest parameters and FLOPs. In contrast, our method achieves the optimal balance between the enhancement performance and the computational complexity of parameters and FLOPs.

4.5 Ablation study

To validate the effectiveness of the proposed network, we perform ablation experiments about the network structure on the LOL dataset.

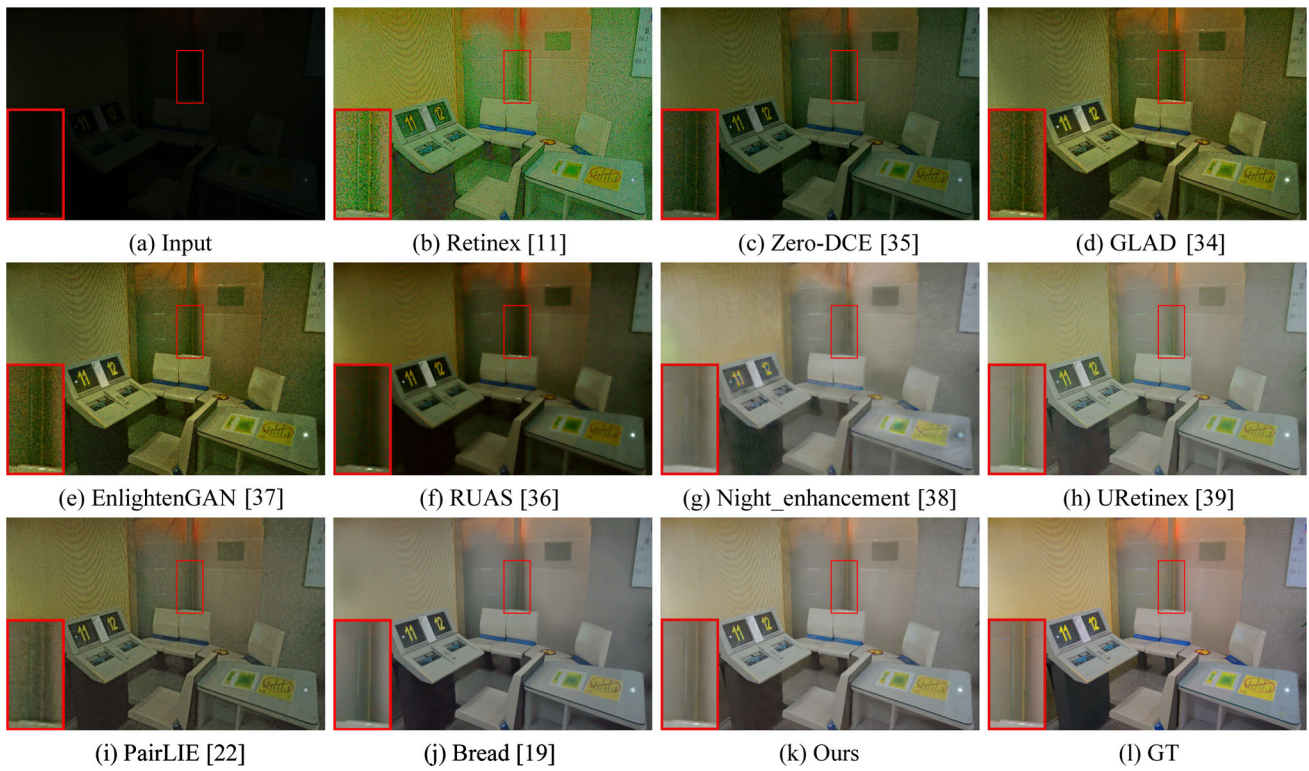


Fig. 4 Visual comparison of all comparison methods on LOL dataset

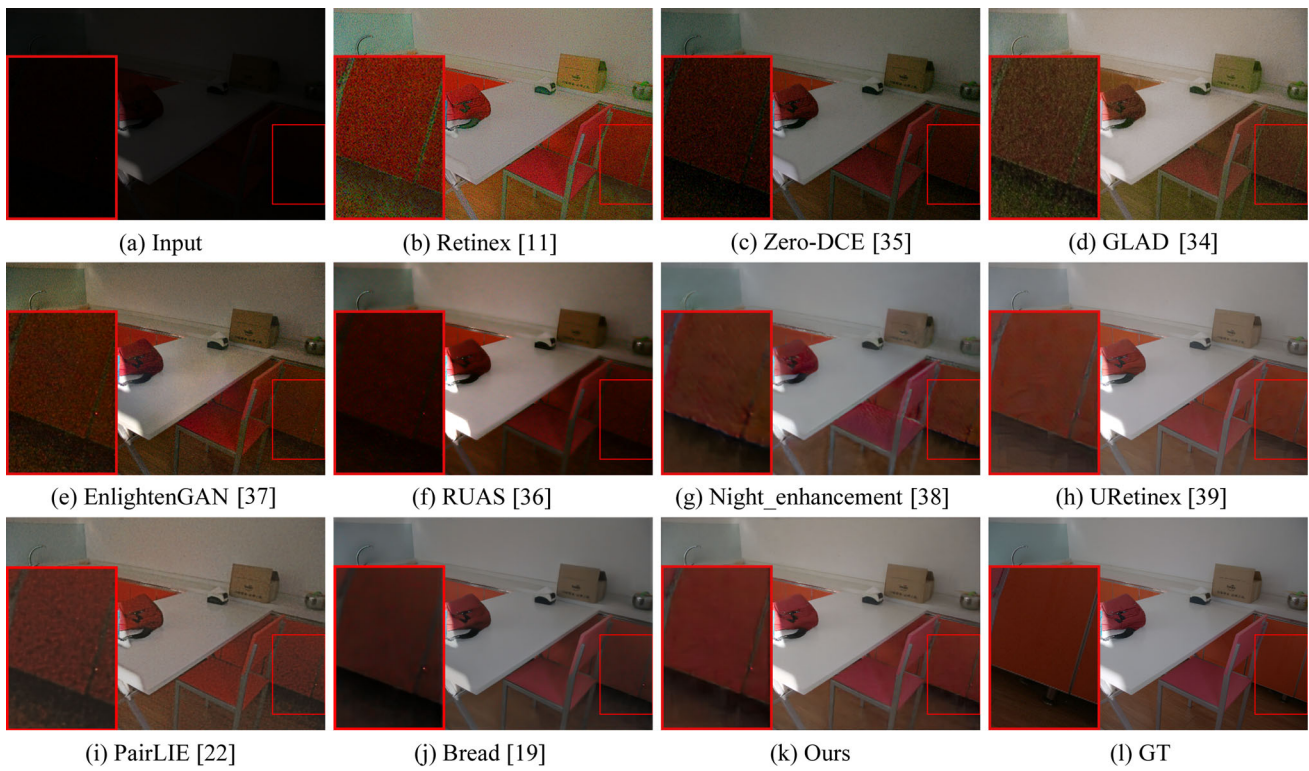


Fig. 5 Visual comparison of all comparison methods on LOL-v2 dataset

Table 2 Complexity comparison of parameters and FLOPs

Methods	Parameters (M)	FLOPs (G)
RetinexNet [11]	0.84	136.02
GLAD [34]	1.13	252.14
Zero-DCE [35]	0.08	19.01
EnlightenGAN [37]	8.64	61.03
RUAS [36]	0.003	0.78
URetinex [39]	0.34	208.50
Night-enhancement [38]	71.73	292.61
PairLIE [22]	0.34	81.84
Bread [19]	2.12	107.07
Ours	2.63	120.66

Table 3 The ablation results of the network structure and the number of IAMs

	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
(a). Ablation on the network structure			
w/o IAM	18.03	0.755	0.271
w/o F_i^i	22.57	0.835	0.132
w/ \bar{L}	23.26	0.840	0.132
w/o IEN	22.97	0.843	0.130
Ours	23.41	0.846	0.129
(b). Ablation on the number of IAMs			
1-1-1-1-1	22.39	0.833	0.139
1-2-3-2-1	22.81	0.839	0.132
2-2-2-2-2	23.22	0.842	0.139
1-2-2-2-1	22.92	0.843	0.132
1-2-2-2-2 (ours)	23.41	0.846	0.129

4.5.1 Components in network structure

To verify the role of each component in the network, four ablation experiments are designed by individually removing or replacing different components in the framework. The results are shown in Table 3a. The introduction of the four ablation experiments is as follows:

- (1) “w/o IAM” represents IACN without IAMs, denoting the IAMs are replaced by ordinary convolutional layers.
- (2) “w/o F_i^i ” represents IACN without the illumination feature maps F_i^i .
- (3) “w/ \bar{L} ” represents replacing L_0 with the output \bar{L} of IEN.
- (4) “w/o IEN” means that the low-light image is used directly as the input of the IACN.

From Table 3a, it can be observed that due to IAM being the core module in the network, its replacement by convolutional layers results in a significant decrease in the metric values obtained by the network. After removing or replacing other components, the performance of the network has decreased to varying degrees. Therefore, the constructed components in our network are effective, which further proves that the two-stage network constructed can achieve good enhancement results.

4.5.2 Number of IAMs

To verify the impact of the number of IAMs on network performance, we test the proposed network by changing the number of IAMs in each scale layer. The results are shown in Table 3b. From the table, it can be seen that increasing the number of IAMs in layers with smaller-scale feature maps does not improve the performance of the model (such as results in the second row). The selection in this article is the optimal among different quantity combinations.

4.6 Evaluation via downstream vision tasks

To evaluate the effectiveness of the proposed method in improving the performance of subsequent visual tasks, an object detection experiment is conducted on the enhanced results of all comparison methods, as shown in Fig. 6. The figure shows the pedestrian detection results using pretrained YOLOv3 on all enhanced images obtained by comparison methods. We can clearly see that more pedestrians can be accurately detected in the result of our method. Therefore, the proposed network achieved enhancement results with higher contrast, which is helpful for improving the performance of subsequent tasks.

4.7 Limitations

Although our method can obtain promising results in most cases, it still has a few limitations. First, as shown in Fig. 7b, our method results in overexposure when the image has nonuniform lighting conditions, as shown in Fig. 7a. Another limitation of our method is that when the input image is captured in an extremely low-light environment (as shown in Fig. 7c), there are noises in our result, as shown in Fig. 7d. In the future, we will design more effective networks to enhance the generalization ability of the models for images under different light distributions.

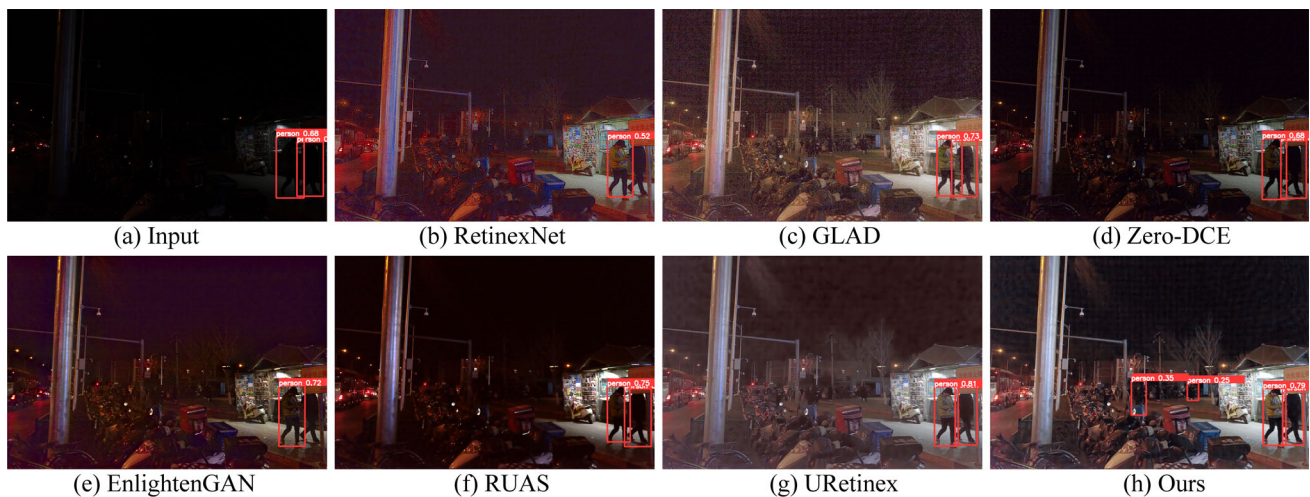


Fig. 6 Visual comparison of object detection results on the enhancement images

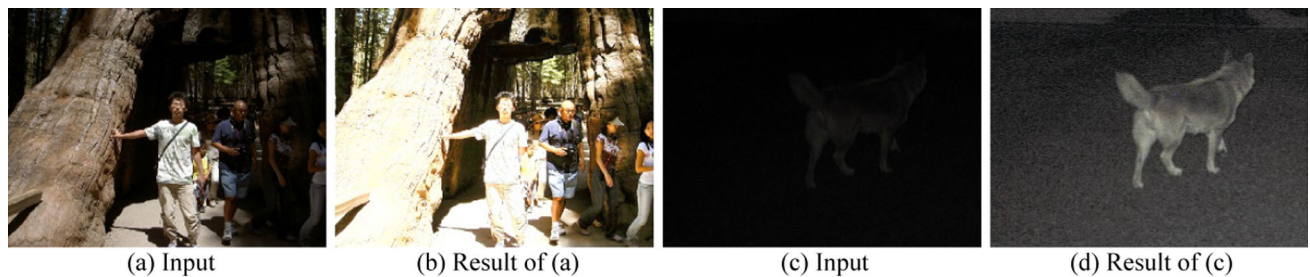


Fig. 7 Failure cases. Input image **a** has nonuniform lighting conditions and **c** is an extremely low-light image. **b** and **d** are the results obtained by the proposed method for images **a** and **c**, respectively

5 Conclusion

In this paper, we propose an IATN based on Retinex theory consisting of two networks: IEN and IACN. The IEN in the first stage is constructed to obtain a preliminary coarse enhancement result by estimating the illumination of the input low-light image. To reduce color distortion and suppress noise in the results obtained by the first stage, IACN in the second stage is constructed as a U-shaped network containing multiply core module IAMs. The IAM is designed to restore image features by establishing long-range dependences of features and the correlation between the image features and illumination features. Extensive experiments on public benchmarks datasets show that our method outperforms some SOTA methods significantly. In future, we will investigate more efficient networks to improve the naturalness and color consistency of LLIE.

Author contributions SH and HD conducted the experiments and wrote part of the paper and helped in methodology, software, writing—original draft. YY revised the paper and was involved in supervision, formal analysis, writing—review & editing. YW and MR conducted the experiments and helped in data curation, software. SW revised the paper and contributed to writing—review & editing.

Funding This work is supported by the National Natural Science Foundation of China (No.62072218 and No.61862030), by the Natural Science Foundation of Jiangxi Province (No.20192ACB20002 and No.20192ACBL21008) and by the Talent project of Jiangxi Thousand Talents Program (No. jxsq2019201056).

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflicts of interest The authors declare no conflict of interest.

References

- Deng, J., Guo, J., Xue, N., Zafeiriou, S.: ArcFace: additive angular margin loss for deep face recognition. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp. 4685–4694 (2019)
- Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2017)
- Fu, J., et al.: Dual attention network for scene segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp. 3141–3149 (2019)
- Abdullah-Al-Wadud, M., Kabir, M.H., Akber Dewan, M.A., Chae, O.: A dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consumer Electron.* **53**(2), 593–600 (2007)
- Wang, L., Xiao, L., Liu, H., Wei, Z.: Variational Bayesian method for retinex. *IEEE Trans. Image Process.* **23**(8), 3381–3396 (2014)
- Guo, X., Li, Y., Ling, H.: LIME: low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **26**(2), 982–993 (2017)
- Moran, S., Marza, P., McDonagh, S., Parisot, S., Slabaugh, G.: DeepLPE: deep local parametric filters for image enhancement. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 12823–12832 (2020)
- Xu, K., Yang, X., Yin, B., Lau, R.W.H.: Learning to restore low-light images via decomposition-and-enhancement. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 2278–2287 (2020)
- 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), New Orleans, LA, USA, pp. 17693–17703 (2022)
- Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: From fidelity to perceptual quality: a semi-supervised approach for low-light image enhancement. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 3060–3069 (2020)
- Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: *BMVC* (2018)
- Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Trans. Image Process.* **30**, 2072–2086 (2021)
- Zhang, Y., Guo, X., Ma, J., Liu, W., Zhang, J.: Beyond brightening low-light images. *Int. J. Comput. Vision* **129**, 1013–1037 (2021)
- Celik, T., Tjahjadi, T.: Contextual and variational contrast enhancement. *IEEE Trans. Image Process.* **20**(12), 3431–3441 (2011)
- Lee, C., Lee, C., Kim, C.-S.: Contrast enhancement based on layered difference representation of 2D histograms. *IEEE Trans. Image Process.* **22**(12), 5372–5384 (2013)
- 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), Las Vegas, NV, USA, pp. 2782–2790 (2016)
- Hao, S., Han, X., Guo, Y., Xu, X., Wang, M.: Low-light image enhancement with semi-decoupled decomposition. *IEEE Trans. Multimed.* **22**(12), 3025–3038 (2020)
- 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, pp. 1632–1640 (2019)
- Guo, X., Hu, Q.: Low-light image enhancement via breaking down the darkness. *Int. J. Comput. Vis.* **131**, 48–66 (2023)
- Liang, Y., Wang, B., Ren, W., Liu, J., Wang, W., Zuo, W.: Learning hierarchical dynamics with spatial adjacency for image enhancement. In: Proceedings of the 30th ACM International Conference on Multimedia, pp. 2767–2776 (2022)
- Yang, Y., Xu, W., Huang, S., Wan, W.: Low-light image enhancement network based on multi-scale feature complementation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 3, pp. 3214–3221 (2023)
- Fu, Z., Yang, Y., Tu, X., Huang, Y., Ding, X., Ma, K.-K.: Learning a simple low-light image enhancer from paired low-light instances. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, pp. 22252–22261 (2023)
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), pp. 6000–6010 (2017)
- Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, pp. 9992–10002 (2021)
- Wang, W., et al.: Pyramid vision transformer: a versatile backbone for dense prediction without convolutions. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, pp. 548–558 (2021)
- Xia, Z., Pan, X., Song, S., Li, L.E., Huang, G.: Vision transformer with deformable attention. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, pp. 4784–4793 (2022)
- Zhang, J., Huang, J., Yao, M., Zhou, M., Zhao, F.: Structure- and texture-aware learning for low-light image enhancement. In: Proceedings of the 30th ACM International Conference on Multimedia (MM '22), pp. 6483–6492 (2022)
- Chan, S.H., Khoshabeh, R., Gibson, K.B., Gill, P.E., Nguyen, T.Q.: An augmented Lagrangian method for video restoration. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, pp. 941–944 (2011)
- Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.: Restormer: efficient transformer for high-resolution image restoration. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, pp. 5718–5729 (2022)
- Lai, W.-S., Huang, J.-B., Ahuja, N., Yang, M.-H.: Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(11), 2599–2613 (2019)
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
- Loshchilov, I., Hutter, F.: Sgdr: stochastic gradient descent with warm restarts. In: *ICLR* (2017)
- Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 586–595 (2018)
- Wang, W., Wei, C., Yang, W., Liu, J.: GLADNet: low-light enhancement network with global awareness. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, pp. 751–755 (2018)
- 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), Seattle, WA, USA, pp. 1777–1786 (2020)

36. Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp. 10556–10565 (2021)
37. Jiang, Y., et al.: EnlightenGAN: deep light enhancement without paired supervision. *IEEE Trans. Image Process.* **30**, 2340–2349 (2021)
38. Jin, Y., Yang, W., Tan, R.T.: Unsupervised night image enhancement: when layer decomposition meets light-effects suppression. In: Proceedings of the European Conference on Computer Vision, pp. 401–421 (2022)
39. Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., Jiang, J.: URetinex-Net: retinex-based deep unfolding network for low-light image enhancement. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, pp. 5891–5900 (2022)

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

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