AMBCR: Low-light image enhancement via attention guided multi-branch construction and Retinex theory

Miao Li | Dongming Zhou | Rencan Nie | Shidong Xie | Yanyu Liu

School of Information Science and Engineering, Yunnan University, Kunming, China

Correspondence
Dongming Zhou, School of Information Science and Engineering, Yunnan University, Kunming 650504, China.
Email: zhoudm@ynu.edu.cn

FUNDING
This work is supported by the National Natural Science Foundation of China (Nos. 62066047, 61966037, 61463052).

Abstract
Due to different lighting environments and equipment limitations, low-light images have high noise, low contrast and unobvious colours. The main purpose of low-light image enhancement is to preserve the details and suppress noise as much as possible while improving the contrast of the image. Here, different networks are first combined to construct a multi-branch module for features extraction, and use the module and Retinex theory to extract the reflection map of the image. Then an attention mechanism is introduced into the multi-branch construction to balance the feature weight of each branch, and get the final result by the reconstruction module. The Retinex theory is used to calculate the $L_1$ loss and the gradient loss for the intermediate feature map of the entire model to train our framework. The entire process is completed in an end-to-end-way, which avoids the hand-crafted reconstruction rules and reduces the workload. What’s more, a large number of experiments demonstrate that the proposed framework performs better results than state-of-the-art algorithms in both quantitative and qualitative evaluations of image enhancement.

KEYWORDS
attention networks, low-light image enhancement, multi-branch construction, Retinex theory

1 | INTRODUCTION

Image enhancement [1–4] is a major research hotspot of interest within the various computer vision tasks [5, 6]. Very often, partial darkness or even global darkness are likely to exist in photos taken by the camera for some reasons, such as weak light, low exposure and backlight. It is difficult to take high-quality images in an environment with insufficient light, although we can adjust the flash and exposure time of the camera to increase the brightness, and the disadvantages of obtaining images via the operations are also unacceptable. For instance, long-time exposure will cause blurred edges, flashlight will introduce uneven shadows and overexposure. It is crucial to obtain clear images in some artificial intelligence tasks such as face recognition [7, 8], target detection [9–11], image fusion [12] and autonomous driving [11, 13].

In addition, denoising is also an important step to be considered for the low-light image enhancement [14, 15]. Many general methods of image enhancement could inevitably amplify noise while increasing brightness [16–20]. In recent years, a large number of models have focused on improving the contrast and brightness of the images [19–24], and it also has a good effect on colour restoration. The existing methods are mainly divided into histogram equalization-based [19, 20, 24] methods and Retinex theory [18, 25–27]. The main goal of the former is to enhance the contrast of the image, and the latter focuses on estimating the illumination component in order to obtain the reflection component of the image. Although significant effects have been made, there are still many weak points that need to be improved. For recognition problem, say contrast enhancement, the details are still not well distinguished due to the amplification of noise even though the brightness is increased. With the improvement of camera equipment, it is more and more important to get a clear image under a certain degree of contrast for our visual experience. However, the degree of contrast is not easy to estimate.

The methods based on deep learning have made great achievements in computer vision tasks [18, 27–29]. Especially in
the tasks of semantic segmentation [30], super-resolution reconstruction [31], target recognition [9] and image denoising [32], more and more excellent models have shown top-level performance. In the field of low-light image enhancement, the existing methods often use a single model to extract feature information; Models based on Retinex theory [33] are usually designed as non-end-to-end models.

Most methods based on Retinex theory used an independent network to decompose the low-light image, and the algorithms of deep learning are often non-end-to-end. Deep UPE focuses on finding the mapping relationship between low-light images and enhanced images and obtains the enhanced image by multiplying the low-light image by the inverse of the illumination mapping, so as to remove the low-level illumination information from the low-illumination image. In contrast, considering the complexity of the image content, we proposed an end-to-end network via attention guided multi-branch construction and Retinex theory (AMBCR) to enhance brightness and preserve the details of a low-light image. In particular, the structure is an end-to-end network, which makes training more convenient.

In this paper, the network sends features in different layers of the image to different branches. Meanwhile, we introduce an attention mechanism to guide the fusion strategy of each branch, and obtain the illumination map containing low-frequency information and the reflection map with detailed information. Next, we fuse feature information to enhance the image. In this manner, the AMBCR is capable of improving the contrast and details of the image, and getting an excellent result.

This study proposes a feasible framework of low-light image enhancement. Our codes will be open. Our main contributions can be summarized as follows.

1. We designed a novel network of low-light image enhancement, which is constructed as an end-to-end model. We use hybrid datasets to train a more robust model. The performance of the model is improved by calculating the loss of the middle layer.
2. We proposed a multi-branch network combined with attention mechanism, and different features are extracted from different network. Then the model reconstructs the features from different branches according to different weights. Thus, the proposed method gets superior performance compared to other algorithms.
3. By combining the Retinex theory and multi-branch network, the proposed method is compared with many state-of-the-art methods via comprehensive experiments, and proved the effectiveness of the methods.

2 RELATED WORKS

Low-light image enhancement has been the main focused area in the field of computer vision. In this section, we mainly introduce some similar methods.

2.1 Conventional methods

The traditional methods are mainly based on two aspects. The first aspect is based on histogram equalization (HE). This method is mainly to improve the image contrast. Ibrahim and Kong [19] enhanced image contrast with dynamic histogram equalization; Pizer et al. [20] optimized the local contrast of the low-light image, it divided the image into each sub-region, and then histogram equalization was performed on each region. Pisano et al. [24] could effectively reduce the noise by limiting contrast. Other traditional methods are based on Retinex theory. The Retinex theory is mainly used to estimate the illumination map of the images. Single-Scale Retinex (SSR) [34] tried to calculate the convolution of the source image and Gaussian function to estimate the illumination map, and had made a breakthrough in the field of low-light image enhancement; Multi-Scale Retinex (MSR) [25] extended the single-scale algorithm to a multi-scale field based on SSR; Multi-Scale Retinex with Colour Restoration (MSRCR) [25] introduced a colour restoration factor to solve the problem of colour distortion which was caused by local enhancement of contrast. Wang et al. [35] balanced the two tasks of detail restoration and naturalness preservation, and it prevented the influence of over-exposure; Fu et al. [36] blended the superiority from different techniques, and designed a multi-scale fusion model for image enhancement; Fu et al. [37] decomposed the image into a reflection map and illumination map via a weighted variational network. In addition, Guo et al. [17] proposed a model which estimated the illumination map according to the maximum value of the corresponding pixel in each channel of the image; Ying et al. [22] designed a multi-exposure fusion framework to achieve the task of low-light image enhancement. In general, traditional methods had enhanced the contrast of image, and restored the colour, but the details of the image were not well restored. They had obvious shortcomings in eliminating noise.

2.2 Deep learning-based methods

With the advent of neural network, a large number of tasks in the field of computer vision have been greatly developed, such as Refs. [31, 38] for super-resolution, Tian et al. [32] for image denoising and Hou et al. [39] for infrared and visible image fusion, etc. These methods have achieved excellent results. LLNet [40] designed a novel stacked-sparse denoising autoencoder for the enhancement of low-light image, the denoising and contrast enhancement are achieved in only one network. In addition, S-LLNet [40] divided the two tasks into two modules, namely, the denoising module and the contrast enhancement module. LLCNN [41] proposed a novel deep convolution module containing inception part [42] and residual part [43] to avoid gradient vanishing problem. MSR-Net [26] combined convolutional neural network (CNN) and Multi-Scale Retinex (MSR) [25] to achieve image enhancement. Based on Retinex Theory [33], Retinex-Net [18] separated a low-light image into two maps which are named reflectance map and illumination map, and then enhanced the illumination map.

---

1 The link of the codes is https://github.com/limiaoair/AMBCR.
LightenNet [44] estimated an illumination part from a non-end-to-end network.

In recent years, Lv et al. [45] established a multi-branch framework named MBLLEN for image enhancement. Chen et al. [46] introduced and used a large number of raw short-exposure low-light images to design an effective deep CNN framework. SICE [23] designed a model which trained multi-exposure images to enhance a single low-light image. Based on RetinexNet, Zhang et al. [27] further designed a low-illumination image enhancement network named KinD that could artificially adjust the brightness, and there are three modules in KinD instead of an end-to-end model. Ren et al. [47] proposed a deep Hybrid network to restore the edge information. Wang et al. [48] designed an encoder network which could extract local features and global features, and the method which is named DeepUPE mainly estimated the illumination map in the middle layer of the network. Jiang et al. [29] established a colour-based attention model for low-light image enhancement. In the field of unsupervised learning, Enlighten-GAN [28] incorporated generative adversarial network (GAN) [49] with low-light image enhancement to solve the problem of image enhancement. Lee et al. [50] proposed an unsupervised approach using bright channel prior theory. Zero-DCE [21] estimated pixel-wise and high-order curves to restore the low-light images. Zhang et al. [51] proposed a self-supervised model and only used one low-light image to complete the training process.

2.3 | Low-light image datasets

Since the dataset has been widely used for deep learning, many famous datasets have been established. For example, the PASCAL VOC [52] contains a large number of image objects, and the ImageNet [53] is widely used for image classification and recognition. However, low-light image dataset is relatively lacking, it is very important to construct a corresponding image dataset in the field of low-light image enhancement. There are two main methods to obtain the low-light image dataset. The first method simulated low-light image by processing the source images [26, 40, 41]. The other is to build an additional dataset, VV-datasets [54, 55] and LIME [17] have dozens of low-light images. Wei et al. [18] established a dataset with large scale paired low-light/normal-light images named LOL-dataset and the dataset is the first to be shot using real scenes. ExDark [56] is the largest low-light image dataset for object focused works which contains annotation. Chen et al. [46] collected a large number of raw images by cameras. The limitation of a single dataset is that the scenario is too simple. In this paper, we have mixed the different datasets to improve the robustness of the model. Experiments show that our model has strong generalization.

3 | METHODOLOGY

In this section, we introduce the proposed network with all the required details, and establish a novel end-to-end network architecture to solve the problem of low-light image enhancement, and perform the decomposed module using the Retinex theory. In addition, our reconstruction module is also included in the AMBCR.

3.1 | Network architecture

We design the AMBCR with the Retinex theory and multi-branch network. It separates the image into sub-feature related to different feature layers. In the network, a low-illumination colour image is inputted, and the output is a clean image with the same size as the input image. The AMBCR architecture is illustrated in Figure 1, and consists of two parts: decomposed net and reconstruction net. In the next subsections, we shall describe them in detail.

3.1.1 | Decomposed net

The decomposed method is mainly based on Retinex theory. In Figure 1, the low-light image is denoted as L-input, and the image is input into the dual channel simultaneously. The two channels are the reflection map extraction module (REM) and the illumination map extraction module (IEM). Inspired by the multi-branch network of MBLLEN [45], the REM attempts to design an attention guided module.

REM

The first part of the REM, there is a single stream network which has 6 convolutional layers, and except for the first two convolutional layers, the input and output features in the remaining four convolutional layers have the same size. Each of the convolutional layers uses kernels of the size 3×3 and ReLU [57] nonlinearity. The sub-features of the last four layers are used as the input of the multi-branch network. The second part of REM includes a multi-branch network and two attention modules. The multi-branch network contains four sub-nets which correspond to the above mentioned four sub-features. The sub-nets consist of two U-net [30], a residual network and a dilated convolutional network. At the front and the rear end of the multi-branch network, there are two attention modules which consist of four SE-nets corresponding to four sub-nets.

IEM

The IEM is used to extract the illumination map, which is composed of a series of convolutional layers. The first three convolutional layers are used to extract the input image into the feature with kernel numbers of 32. After connecting the output feature of REM, a single-channel output is obtained by another three convolutional layers of which the last layer uses sigmoid nonlinearity.

3.1.2 | Reconstruction net

The reconstruction part is mainly composed of convolutional layers. It accepts the outputs of REM and IEM to obtain the
enhanced image. We first concatenate them in colour channel dimension, and then use convolutional layers with kernel numbers of 16, 16 and 3 to extract their features respectively. Instead of ReLU, we still choose Sigmoid as the activation function in the last layer.

### 3.2 Loss function

In this section, we design the loss function which aims to achieve the work of the low-light image enhancement and find suitable parameters to sufficiently express the coefficient of the loss function.

We plan to design a robust loss function, but in this case, it is insufficient for our network to use the simple mean square error (MSE). Therefore, we establish a novel compound loss function.

The loss $L$-sum consists of two parts: adversarial decomposition loss and reconstruction loss:

$$L_{\text{sum}} = L_{\text{recon}} + \lambda L_{\text{decom}},$$  \hspace{1cm} (1)

where $\lambda$ denotes the coefficient to adjust the impact of decomposition loss on result, $L_{\text{decom}}$ uses the intermediate feature map of the model to calculate the loss and it makes the network as far as possible to get a better reflection map based on the Retinex theory. On the other hand, $L_{\text{recon}}$ can always make the model reconstructing and image enhancing more preferably.

#### 3.2.1 Reconstruction loss

We build the reconstruction loss function as the main loss of the model, and it contains three parts: gradient loss, SSIM $[16]$ loss and MSE loss. The objective is as follows:

$$L_{\text{recon}} = L_{\text{mse}} + L_{\text{ssim}} + L_{\text{gradient}},$$  \hspace{1cm} (2)

where, MSE loss is the $l_2$ norm, and SSIM loss should be able to express the difference in structure, the loss function of the first two parts can be expressed as:

$$L_{\text{loss}_{\text{mse,ssim}}} = \left\| \hat{L} - L \right\|_2^2 + \left[ 1 - \frac{1}{3} \sum_{k=1}^{3} \text{ssim} \left( L^k, \hat{L}^k_{\text{high}} \right) \right],$$  \hspace{1cm} (3)

where, the subscript $k$ represents the different colour channels of the image. $L^k$ represents the channel with number $k$ of the reflection map of the enhanced image, $\hat{L}^k_{\text{high}}$ denotes the channel with number $k$ of the reflection map of the
high-light image. The SSIM is a structural loss function, which compares the brightness, contrast, and structure of the two images to obtain an objective evaluation standard. The formula is as follows:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\mu_x^2 + \mu_y^2 + C_1(\sigma_x^2 + \sigma_y^2) + C_2},$$  \hspace{1cm} (4)$$

where, \(\mu\), \(\sigma\) and \(\sigma_{xy}\) denote the mean, standard deviation and correlation between \(x\) and \(y\), respectively. In addition, the third part is gradient loss, which is mainly used for restoring details of the image. The loss function of gradient is as follows.

$$L_{\text{gradient}} = \left[ (\nabla I_{\text{x}} - \nabla I_{\text{y}})^2 + (\nabla I_{\text{y}} - \nabla I_{\text{x}})^2 \right] / 2,$$  \hspace{1cm} (5)$$

where \(\nabla I_{\text{x}}, \nabla I_{\text{y}}\) are the gradients in the \(x\) and \(y\) dimension of the grayscale of the enhanced image, similarly \(\nabla I_{\text{x}}, \nabla I_{\text{y}}\) are the gradient of the label. \(\nabla\) is the first order derivative operator, and the subscripts represent the horizontal and vertical directions respectively.

### 3.2.2 Decomposition loss

Since the purpose of the decomposition loss is to estimate the illumination map via Retinex theory, the focus of the loss function is to get a better illumination map. Therefore, we build a composite loss function which was mainly designed for obtaining the content of the illumination map. The loss function is still composed of three parts:

$$L_{\text{mean}} = L_{\text{Retinex}} + \beta L_{\text{gradient}}^1 + \gamma L_{\text{gradient}}^2,$$  \hspace{1cm} (6)$$

where, the purpose of \(L_{\text{Retinex}}\) is to make the obtained illumination and reflection map conform to the theoretical results of Retinex theory as much as possible. The mathematical model of the Retinex theory is formulated as:

$$L(x, y) = R(x, y)gI(x, y),$$  \hspace{1cm} (7)$$

where, \(R(x, y)\) is a reflection map and \(I(x, y)\) is an illumination map. As a result, the loss function turns out to be: \(L_{\text{Retinex}} = RI - I_{\text{input}}\). In addition, in order to improve the effect of the illumination map, we use the last two parts of the loss function to penalize the illumination map. The gradient loss is the illumination smoothness loss, which is constrained as follows:

$$L_{\text{gradient}}^1 = \mu \left\| \frac{\nabla L_x}{\max(\|\nabla L_x\|, \epsilon)} \right\|_1 + \tau \left\| \frac{\nabla L_y}{\max(\|\nabla L_y\|, \epsilon)} \right\|_1,$$  \hspace{1cm} (8)$$

where, \(\cdot\) means the \(L_1\) loss, and the first part of the gradient loss is the gradient loss between the illumination maps which obtained by IEM module by the low-light image and the high-light image respectively (see Figure 2).

$$L_{\text{gradient}}^2 = \mu \left\| \frac{\nabla R_x}{\max(\|\nabla L_x\|, \epsilon)} \right\|_1 + \tau \left\| \frac{\nabla R_y}{\max(\|\nabla L_y\|, \epsilon)} \right\|_1.$$  \hspace{1cm} (9)$$

Subsequently, the gradient loss of the second part compares the relationship between the reflection map components of different inputs, which are the low-light image and high-light image. The specific relationship is shown in Figure 2.

### 4 EXPERIMENTS

In this part, a large number of experiments are used to verify the effectiveness of the AMBCR-net. Moreover, we not only use the gradient loss between the illumination map and the input image to validate the model. Next, we demonstrate the results of comparison between the proposed network and the state-of-the-arts methods.

#### 4.1 Implementation details

We use hybrid datasets as the training dataset, which includes LOL-dataset and SICE-dataset. In the training, we employ 480 image pairs of LOL-dataset, and 20 image pairs of SICE-dataset respectively. The training images are only cropped to a uniform size without additional artificial operations. In general, we use a total of 500 pairs of images as the training dataset. For the proposed network, the batch size is set to be 2 and the patch size to be 256x256. We set the parameters as: \(\alpha = 0.1, \beta = 0.2, \lambda = \mu = 1\). Our network was trained for 350 epochs, and the model was minimized by the Adam optimizer which has a learning rate of \(10^{-4}\). The AMBCR was implemented on Tensorflow 1.15 framework [58] and the entire model was trained on Nvidia GTX2080Ti GPU, Intel Core i7-9700KF 3.6 GHz CPU and 32G RAM.
4.2 Performance evaluation and contrastive analysis

4.2.1 Qualitative results

In this section, lots of experiments are used to verify the effects of our model. Our enhancement framework is compared with the traditional methods and the deep learning methods. The traditional methods include Dong [59], Illumination Map Estimation (LIME) [17], Exposure Fusion Framework (BIMEF) [22] and Naturalness Preserved Enhancement Algorithm (NPE) [35]. The deep learning methods include RetinexNet [18], KinD [27], MBLLEN [45] and a curve estimation model based on zero-reference (Zero-DCE) [21]. Moreover, we used two public datasets, which include LOL-dataset [18] and SICE-dataset [23], and the codes of the baseline algorithms are from the original authors. We first show the results of the experiment, and then compare the subjective and objective analysis.

The enhanced image is initially shown in Figure 3. We concentrate on the contrast, the clarity of detail and the degree of colour restoration of each result. Because the brightness is more dependent on the perception of the supervisor, the image brightness is still lower for some labels of LOL datasets. Our method enhanced the brightness of the low-light image well, and got a larger improvement in the visualization than the low-light image. Figure 4 is the histogram of Figure 3.

In addition to contrast enhancement, the results are shown in Figure 5. Figure 5(j) shows the normal-light image which is used as a comparison of the details, detail parts used for comparison are denoted with the white and red dotted box. As the method of Dong [59] loses some details due to only processing the reversed image by using the dehazing method, the result is shown in Figure 5(a).

And the result of the MF [36] is better than the Dong [59], but the details are still fuzzy. Although optimized the bright channel part, the results of LIME [17] and NPE [35] still tend to be darker and blurry, for instance, the contour details of plant in the white box are a bit unclear. In contrast, the result by BIMEF [22] has better details by using a multi-exposure fusion framework, but the enhancement of brightness is still not enough. We note that the results obtained by KinD [27] and MELLEN [45] reduce the impact of noise. The main reason is that these two methods use neural networks for supervised learning. But the details of the MBLLEN [45] are also still unclear, and our results are brighter than KinD [27].

In the “Toy” image in Figure 6, we found that in the magnified image marked by the white frame, our result has a clearer texture than other methods. The texture of LIME is also relatively clear, but there is a lot of noise in the result of LIME. Compared with other deep learning methods, our method can improve visibility and brightness while preserving the colour of the image.

What’s more, a more detailed comparison experiment is shown in Figure 7. And we selected five images for comparison. The low-light and high-light images are displayed in the first and last row of the results in Figure 7. The results show in Figure 7 that our model not only is better than other classic algorithms, but also has excellent visual and contrast effects compared with advanced models of deep learning such as KinD [27] and MBLLEN [45].

We also conducted user studies (US) to evaluate the visual quality of enhanced images obtained by different methods. We selected 20 participants to rate enhanced images using
FIGURE 4  Grayscale histogram corresponding to the image in Figure 3. The first row shows the grayscale histogram of “Swimming pool”, the second rows shows the grayscale histogram of “Toy”.

FIGURE 5  Visual comparison with state-of-the-art image enhancement methods in “Swimming pool”
different methods. 80% of the participants are engaged in computer vision research, and the remaining 4 participants are engaged in computer science and automation science and other related industries. We mainly grade according to the following three points: (1) whether there exists obvious colour deviation; (2) whether the results lead to overexposure or underexposure; (3) whether the noise is newly generated. The scores of user study range from 1 to 5, and the higher the score, the better the subjective quality of the enhanced images is. We scored all the test images in the LOL dataset and the average scores of user study for LOL-dataset are shown in Table 5.

4.2.2 Quantitative results

The subjective visual evaluation is easily too affected by personal emotion. For individuals, identifying the specific details and potential noise of the enhanced image is extremely difficult. Therefore, it is indispensable using quantitative evaluations. We utilize five reliable quantitative metrics, including peak signal to noise ratio (PSNR), signal to noise ratio (SNR), structural similarity (SSIM), mean square error (MSE), visual information (VIF) [60] and natural image quality evaluator (NIQE) [61]. The first three indexes are evaluation indicators with reference, which need to use the high-light image as labels, while NIQE is a non-reference evaluation index and only a single test image is required. We perform quantitative analysis on the test images, as shown in Tables 1–3.

Table 1 reports the performance of the enhanced results of nine methods including our algorithm on the five images in Figure 6. From the numerical results, we can see that most of the indexes of our AMBCR are significantly better than the

| Methods    | PSNR | SSIM | SNR | VIF | MSE   | NIQE |
|------------|------|------|-----|-----|-------|------|
| Dong [59]  | 20.93| 0.60 | 1.50| 0.51| 525.45| 8.52 |
| Zero-DCE [21] | 15.48| 0.81 | 7.63| 0.45| 1841.73| 7.37 |
| LIME [17]  | 20.79| 0.58 | 4.99| 0.60| 542.18| 8.50 |
| BIMEF [22] | 13.01| 0.72 | 0.25| 0.53| 3252.25| 7.33 |
| NPE [35]   | 16.39| 0.65 | 0.37| 0.56| 1491.42| 8.29 |
| RetinexNet [18] | 21.78| 0.59 | 3.02| 0.55| 431.16| 10.06 |
| KinD [27]  | 18.30| 0.81 | 6.12| 0.55| 164.26| 5.39 |

Table 2 reports the performance of the enhanced results of “Data-2”.

| Methods    | PSNR | SSIM | MSE  | VIF  | SNR  | NIQE |
|------------|------|------|------|------|------|------|
| Dong [59]  | 10.90| 0.326| 5286.67| 0.45| 0.05| 9.03 |
| Zero-DCE [21] | 13.40| 0.71 | 3165.05| 0.48| 5.68| 6.85 |
| LIME [17]  | 14.71| 0.30 | 2197.03| 0.41| 1.03| 9.11 |
| BIMEF [22] | 8.91 | 0.40 | 8358.98| 0.47| 0.08| 8.42 |
| NPE [35]   | 13.18| 0.33 | 3128.85| 0.54| 0.32| 9.31 |
| RetinexNet [18] | 13.92| 0.30 | 2636.40| 0.40| 0.98| 7.63 |
| KinD [27]  | 18.75| 0.81 | 867.45| 0.51| 1.34| 5.12 |
| MBLLEN [45] | 20.81| 0.78 | 539.21| 0.43| 0.67| 4.06 |

| Methods    | PSNR | SSIM | MSE  | VIF  | SNR  | NIQE |
|------------|------|------|------|------|------|------|
| Dong [59]  | 10.90| 0.326| 5286.67| 0.45| 0.05| 9.03 |
| Zero-DCE [21] | 13.40| 0.71 | 3165.05| 0.48| 5.68| 6.85 |
| LIME [17]  | 14.71| 0.30 | 2197.03| 0.41| 1.03| 9.11 |
| BIMEF [22] | 8.91 | 0.40 | 8358.98| 0.47| 0.08| 8.42 |
| NPE [35]   | 13.18| 0.33 | 3128.85| 0.54| 0.32| 9.31 |
| RetinexNet [18] | 13.92| 0.30 | 2636.40| 0.40| 0.98| 7.63 |
| KinD [27]  | 18.75| 0.81 | 867.45| 0.51| 1.34| 5.12 |
| MBLLEN [45] | 20.81| 0.78 | 539.21| 0.43| 0.67| 4.06 |
FIGURE 7  Enhanced results of five sets of images obtained by different enhancement methods
algorithms listed, only the values of NIQE [61] are in the second and third place in Tables 1 and 2. Due to space limitations, we list the quantitative indexes of three test images Data-1, Data-2 and Data-3 in Tables 1–3. Because of the exposure of the source images and a perfect image with normal lighting, we use NIQE [61] as a reference. The average performances of the test images are shown in Table 4. The best results are marked in bold.

Table 4 compares the quantitative assessment of different methods using six indexes, PSNR and MSE are unified indicators, KinD [27] and MBLLEN [45] of deep learning methods have better performance. The main reason is that the ability to extract features of neural networks is greatly improved compared to traditional methods. In SSIM, the value closer to one means better enhancement. In addition, we have a huge advantage in SNR, the smaller the value is, the smaller the noise proportion is. What’s more, we have good results in the other listed indexes.

For further analysis, the colour restoration effect of our result is closer to the high-light image in Figure 7. As for the index of contrast, the results of KinD [27] and MBLLEN [45] are some-what the same, but our results are more effective in extreme dark parts of the images. The comprehensive comparison results are shown in Figures 8–9. As shown in Figure 8(c), our method has a higher signal-to-noise ratio (SNR), which has huge advantages compared with other methods. In addition, the number of NIQE [61] also has the same performance as the advanced algorithm.

Figure 9 illustrates the average results of the test images. It can be seen that the deep learning methods have obvious advantages in quantitative indicators of PSNR and SSIM, and our enhanced image has achieved better results. The right side of Figure 9 shows the radar chart of the PSNR with five test images. Our results have good advantages.

Table 5 shows the average performance indicators of all test images for our method and other methods on the LOL-dataset. For the user study (US) score, our method is slightly ahead, and KinD and LIME also have the advantage in visual quality. In addition, we compared the running time. The running time of our method is 0.04 s per image, which is slightly lower than the time of Zero-DCE and meets the requirements of real-time processing.

### 4.3 Extended experiment

We also used the dataset provided by SICE [23] to verify the effectiveness of our model. The SICE-dataset is a multi-exposure image dataset containing 229 sets of multi-exposure images and each set contains labels. We selected the lowest exposure image in each group, and the comparison results are shown in Figure 10.

The results are shown in the first row of Figure 10. The contrast of the enhanced image has been significantly improved, and our results are closer to the label images. In the results of the second and third rows, the enhanced image of BIMEF [22] and Dong [59] have lower contrast, and focus on the part marked by the red box in the third row; details in results of MF [36], Dong [59], BIMEF [22] and LIME [17] looks so blurry that we cannot distinguish, the results of KinD [27] and MBLLEN [45] are closest to our result, but our enhanced image has better quantitative indicators, and results are shown in Table 6. Then, we calculated the difference of the illumination map obtained by the middle layers of our model under different exposure rates.

As shown in Figure 11, we listed the colour-map to show the distribution of the illumination map. It can be seen that the AMBCR is able to effectively extract illumination information under various exposure conditions. With the increase of the exposure intensity, the brightness of the extracted illumination map also increased, which can demonstrate the effectiveness of the intermediate loss $I_{\text{detail}}$.

In the colour-map and illumination map, we can find that the definition of detail gradually increased with the exposure rate. In the problem of low-light image enhancement, the illumination map we obtained is more accurate and contains less information than the reflection map.
FIGURE 8  Quantitative comparisons of the four metrics, that is, PSNR, SSIM, SNR and NIQE on five image sets

TABLE 5  Average quantitative assessment comparison and the US scores of different methods for the LOL-dataset

| Methods     | DONG   | NPE    | LIME   | BIMEF  | RetinexNet | MBILEN | KinD  | Zero-DCE | Proposed |
|-------------|--------|--------|--------|--------|------------|--------|-------|----------|----------|
| PSNR        | 16.20  | 16.78  | 17.18  | 13.05  | 17.82      | 20.38  | 22.28 | 16.42    | 24.75    |
| SSIM        | 0.44   | 0.44   | 0.40   | 0.56   | 0.41       | 0.74   | 0.84  | 0.69     | 0.86     |
| SNR         | 1.02   | 2.05   | 3.78   | 0.21   | 3.60       | 2.85   | 4.62  | 5.65     | 8.03     |
| US          | 2.88   | 2.98   | 3.65   | 2.88   | 2.99       | 3.61   | 3.85  | 3.10     | 3.96     |
| TIME        | 0.065  | 5.030  | 0.102  | 0.091  | 0.048      | 0.056  | 0.051 | 0.035    | 0.041    |
TABLE 6: Average quantitative assessment comparison of different methods for the enhancement results of “Img-1-3″

| Methods   | Img-1 PSNR | Img-1 SSIM | Img-2 PSNR | Img-2 SSIM | Img-3 PSNR | Img-3 SSIM |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| KinD[27]  | 19.89     | 0.87      | 21.63     | 0.68      | 16.83     | 0.70      |
| BIMEF[22] | 17.77     | 0.86      | 12.70     | 0.55      | 11.68     | 0.56      |
| DONG[38]  | 14.85     | 0.74      | 20.34     | 0.66      | 17.00     | 0.69      |
| LIME[17]  | 14.77     | 0.78      | 22.63     | 0.69      | 19.93     | 0.71      |
| MBLLEN[45]| 14.20     | 0.75      | 21.40     | 0.62      | 13.15     | 0.58      |
| NPE[35]   | 19.31     | 0.81      | 20.53     | 0.67      | 16.26     | 0.70      |
| Proposed  | 25.84     | 0.89      | 23.18     | 0.71      | 24.59     | 0.81      |

To better show the performance of the different methods, Figure 12 shows the visualization results of the enhanced images obtained by the different methods. Image details in most regions are revealed, but in order to determine the saturation of the region, we set the binary mask visualization based on an adjustable threshold. First, we convert the enhanced image into a grayscale image, and then adjust the threshold to obtain a binary image. Our purpose is to distinguish between over-saturated and under-saturated regions. On the one hand, for over-saturated regions, the pixel value higher than 200 is assigned as white, otherwise black. On the other hand, for under-saturated regions, the pixel value less than 50 is assigned as black, otherwise white. The results are shown in Figures 13 and 14.

In Figures 13 and 14, we focus on the high-light part and low-light part of the enhanced images. We set the specified pixel of the enhanced images to 1 or 0 by using the form of binary figure, Figures 13 and 14 can see clearly the local highlight areas and dark areas in the image. The edges of enhanced images can be compared well.

4.4 Video enhancement

In order to verify the validity of our model, we added an experiment of video enhancement. We choose GOT-10K [62] dataset and use specific software to reduce the contrast and the brightness of the videos. We tested the model on the low-light video and the results were shown in Figure 15. In Figure 15, we selected the key frames for subjective analysis, and our results have a good effect on enhancing brightness and restoring details.

4.5 Ablation experiment

4.5.1 Parameter selection

In order to ensure the optimal performance of the enhanced model, we need to compare the different performance of the different parameters further. Then, we changed the proportion of the intermediate loss, which was denoted by $\lambda$, we set it to 0, 0.1, 0.2 and 0.3. In addition, we also discuss the role of the attention module in the FEM, and compare the results of our model with the same one without the attention module. We choose the LOL-dataset [18] for analysis in Tables 7-8 and Figure 16.

As shown in the Tables 7-8, our method got better results when $\gamma = 0.1$, and only one image in PSNR is slightly lower than...
**FIGURE 10** Enhanced results of three image sets obtained by different enhancement methods in SCIE-dataset

**FIGURE 11** The enhanced results of multi-exposure images. The first row shows the source images. The second, third and fourth rows are the reflection map, the illumination map and the colourmap of the illumination map, respectively.
FIGURE 12  Visualized results of Image-1 on SICE-Dataset. The first row shows the grayscale image of the source images. The second row shows the colourmap of the same images

FIGURE 13  Binary mask visualization of Figure 14 (Intensity higher than 200 would be assigned 1, otherwise 0)

TABLE 7  Average PSNR of the proposed method with different parameters

| Methods | $\lambda = 0$ | $\lambda = 0.1$ | $\lambda = 0.2$ | $\lambda = 0.3$ |
|---------|--------------|----------------|----------------|----------------|
| Data-1  | 25.75        | 25.98          | 25.09          | 25.07          |
| Data-2  | 26.15        | 27.03          | 26.05          | 25.64          |
| Data-3  | 26.13        | 25.25          | 26.45          | 26.65          |
| Data-4  | 22.30        | 22.36          | 22.10          | 22.11          |
| Data-5  | 24.53        | 25.29          | 23.67          | 24.43          |

TABLE 8  Average SSIM of the proposed method with different parameters

| Methods | $\lambda = 0$ | $\lambda = 0.1$ | $\lambda = 0.2$ | $\lambda = 0.3$ |
|---------|--------------|----------------|----------------|----------------|
| Data-1  | 0.8985       | 0.8998         | 0.8916         | 0.8915         |
| Data-2  | **0.9002**   | 0.8992         | 0.9009         | 0.8986         |
| Data-3  | 0.8349       | **0.8371**     | 0.8384         | 0.8349         |
| Data-4  | 0.7770       | **0.7784**     | 0.7767         | 0.7722         |
| Data-5  | 0.8852       | **0.8888**     | 0.8789         | 0.8820         |
4.5.2 Structure ablation

The results of AMBCR are obviously better than the model without attention module, and Figure 16(a,b,c) shows the results of quantitative indicators under different parameters, the comprehensive performance of $\alpha = 0.1$ is the best in Figure 16(b), and the effect of each image is relatively small.

In addition, we discussed the impact of the different branches on the results. We respectively removed one of the four branches and then trained the model. What’s more, in order to prove that more branches will not bring better results, we doubled the original number of branches to an eight-branch network. The final comparison results and spatial structure of the different experiments are shown in Figure 17. In Figure 17, we can see that the final results of the model, which include four branches, are closer to the high-light images in the spatial structure. Table 9 shows the average indicators of all test images in LOL-dataset [18] using different structures.

| Eight-branches | Remove CNN | Remove Residual | Remove U-net | Proposed |
|----------------|------------|-----------------|--------------|----------|
| PSNR           | 20.86      | 19.43           | 20.48        | 20.26    | 21.62    |
| SSIM           | 0.84       | 0.82            | 0.83         | 0.83     | 0.85     |
| MSE            | 601.45     | 820.21          | 627.65       | 696.78   | 505.91   |
CONCLUSION

In this paper, we design a novel end-to-end network of low-light image enhancement. We combine Retinex theory and multi-branch network for training. Our network consists of three modules, namely, the REM, IEM and reconstruction module. Our method uses a mixed loss function, and especially, we use the intermediate loss to verify the effect of our model. A large number of experiments prove that we can extract illumination maps to enhance the low-light images well. On this basis, we can perform further processing on the enhanced images, such as face recognition, target detection and target tracking. According to the Retinex theory, this model is helpful for multi-exposure images fusion.

ACKNOWLEDGEMENTS

The authors thank the editors and anonymous reviewers for their detailed reviews, valuable comments and constructive suggestions for this study. This work is supported by the National Natural Science Foundation of China (Nos. 62066047, 61966037, 61463052).

ORCID

Miao Li https://orcid.org/0000-0001-6196-2041
FIGURE 17  The structural ablation experiment results. (a1)–(a3) The eight-branches of the REM. (b1)–(b3) Remove CNN module from the REM as the three-branch structure. (c1)–(c3) Remove Residual module from the REM. (d1)–(d3) Remove U-net module from the REM. (e1)–(e3) Results of proposed. (f1)–(f3) High-light images. (a1)–(f1) The enhanced images of the different structure. (a2)–(f2) The grayscale images of the different structure. (a3)–(f3) The spatial structure of the different structure.

REFERENCES
1. Chen, D., et al.: Controllable Image Processing via Adaptive FilterBank Pyramid. IEEE Trans. Image Process. PP(99), 1–1 (2020)
2. Qiu, T., et al.: Efficient medical image enhancement based on CNN-FBB model. IET Image Proc. 13(10), 1736–1744 (2019)
3. Chaoying, T., et al.: Low-light image enhancement with strong light weakening and bright halo suppressing. IET Image Proc. 13(3), 537–542 (2019)
4. Ren, X., et al.: LR3M: Robust low-light enhancement via low-rank regularized retinex model. IEEE Trans. Image Process. 29, 5862–5876 (2020)
5. Liu, Y., et al.: Construction of high dynamic range image based on gradient information transformation. IET Image Proc. 14(7), 1327–1338 (2020)
6. Stark, A.J.: Adaptive image contrast enhancement using generalizations of histogram equalization. IEEE Trans. Image Process. 9(5), 889–896 (2000)
7. Kar, A., et al.: LMZMPM: Local modified zernike moment per-unit mass for robust human face recognition. IEEE Trans. Inf. Forensics Secur. 14(8), 495–509 (2015)
8. Cheng, Z., Zhu, X., Gong, S.: Face re-identification challenge: Are face recognition models good enough? Pattern Recognit. 107, 107422 (2020)
9. Moradi, S., Moallem, P., Sabahi, M.F.: Fast and robust small infrared target detection using absolute directional mean difference algorithm. Signal Process. 177, 107727 (2020)
10. Jung, H., et al.: Unsupervised deep image fusion with structure tensor representations. IEEE Trans. Image Process. 29(99), 3845–3858 (2020)
11. Liu, D., et al.: Video object detection for autonomous driving: Motion-aid feature calibration. Neurocomputing 409, 1–11 (2020)
12. Liu, Y., et al.: Multimodal medical image fusion based on the spectral total variation and local structural patch measurement. Int. J. Imaging Syst. Technol. 31(1), 391–411 (2021)
13. Chen, L., et al.: DenseLightNet: A light-weight vehicle detection network for autonomous driving. IEEE Trans. Ind. Electron. 67(12), 10600–10609 (2020)
14. Kang, E., et al.: Deep convolutional framelet denoising for low-dose CT via wavelet residual network. IEEE Trans. Med. Imaging 37(6), 1358–1369 (2018)
15. You, Q., et al.: Fundus image enhancement method based on cycleGAN. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin (2019)
16. Ma, K., et al.: Multi-exposure image fusion by optimizing a structural similarity index. IEEE Trans. Comput. Imaging 4(1), 60–72 (2018)
17. Guo, X., Li, Y., Ling, H.: LIME: Low-light image enhancement via illumination map estimation. IEEE Trans. Image Process. 26(2), 982–993 (2017)
18. Wei, C., et al.: Deep retinex decomposition for low-light enhancement. The British Machine Vision Conference (BMVC), Newcastle, UK (2018)
19. Ibrahim, H., Kong, N.S.P.: Brightness preserving dynamic histogram equalization for image contrast enhancement. IEEE Trans. Consum. Electron. 53(4), 1752–1758 (2007)
20. Pizer, S.M., et al.: Adaptive histogram equalization and its variations. Comput. Vis. Graphics Image Process. 39(3), 356–368 (1987)
21. Guo, C., et al.: Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, California, pp. 1780–1789 (2020)
22. Ying, Z., et al.: A new image contrast enhancement algorithm using exposure fusion framework. In: International Conference on Computer Analysis of Images and Patterns, Salerno, Italy, pp. 36–46, (2017)
23. Cai, J., et al.: Learning a deep single image contrast enhancer from multi-exposure images. IEEE Trans. Image Process. 27(4), 2049–2062 (2018)
24. Pisano, E.D., et al.: Contrast limited adaptive histogram equalization for image contrast improvement of mammograms. J. Digital Imaging 11(4), 193 (1998)
25. Jobson, D.J., Rahman, Z.U., Woodell, G.A.: A multiscale retinex. IEEE Trans. Image Process. 6(3), 451–462 (1997)
26. Honda, T., Sugimura, D., Hamamoto, T.: Multi-frame RGB/NIR imaging of a center/surround retinex. IEEE Trans. Image Process. 6(7), 965–976 (1997)
27. Shen, L., et al.: Mr-net: Low-light image enhancement using deep convolutional network. arXiv preprint arXiv:1711.02488, (2017)
28. 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, Nice, France pp. 1632–1640 (2019)
29. Jiang, Y., et al.: Enlightengan: Deep light enhancement without paired supervision. arXiv preprint arXiv:1906.06972, (2019)
30. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, Munich, Germany, pp. 234–241 (2015)
31. Honda, T., Sugimura, D., Hamamoto, T.: Multi-frame RGB/NIR imaging for low-light color image super-resolution. IEEE Trans. Comput. Imaging 6, 248–262 (2020)
32. Tian, C., et al.: Attention-guided CNN for image denoising. Neural Networks 124, 117–129 (2020)
33. Land, E.H., McCann, J.J.: Lightness and retinex theory. Josa 61(1), 1–11 (1971)
34. Jobson, D.J., Rahman, Z.U., Woodell, G.A.: Properties and performance of a center/surround retinex. IEEE Trans. Image Process. 6(3), 451–462 (1997)
35. Wang, S., et al.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Trans. Image Process. 22(9), 3538–48 (2013)
36. Fu, X., et al.: A fusion-based enhancing method for weakly illuminated images. Signal Process. 129, 82–96 (2016)
37. Fu, X., et al.: A weighted variational model for simultaneous reflectance and illumination estimation. In: 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada, pp. 2782–2790 (2016)
38. Liu, J., et al.: Residual feature aggregation network for image super-resolution. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle (2020)
39. Hou, R., et al.: VIF-Net: An unsupervised framework for infrared and visible image fusion. IEEE Trans. Comput. Imaging (6), 640–651 (2020)
40. Lore, K.G., Akintayo, A., Sarkar, S.: LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognit. 61, 650–662 (2017)
41. Tao, L., et al.: LLCNN: A convolutional neural network for low-light image enhancement. In: 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, Florida, pp. 1–4, (2017),
42. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, Massachusetts, pp. 1–9 (2015)
43. He, K., et al.: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada, pp. 770–778 (2016)
44. Li, C., et al.: LightenNet: A convolutional neural network for weakly illuminated image enhancement. Pattern Recognit. Lett. 104, 15–22 (2018)
45. Lv, F., et al.: MBILLEN: Low-light image/video enhancement using CNNs. In: The British Machine Vision Conference(BMVC), Newcastle, UK (2018)
46. Shen, L., et al.: Learning to see in the dark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, pp. 3291–3300 (2018)
47. Ren, W., Liu, S., Ma, L., et al.: Low-light image enhancement via a deep hybrid network. IEEE Trans. Image Process. 28(9), 4364–4375 (2019)
48. Wang, R., et al.: Underexposed photo enhancement using deep illumination estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, California, pp. 6849–6857 (2019)
49. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, Montreal, Canada, pp. 2672–2680 (2014)
50. Lee, H., Sohn, K., Min, D.: Unsupervised low-light image enhancement using bright channel prior. IEEE Signal Process. Lett. 27, 251–255 (2020)
51. Zhang, Y., et al.: Self-supervised image enhancement network: Training with low light images only. In: Computer Vision and Pattern Recognition, Seattle (2020)
52. Everingham, M., et al.: The Pascal visual object classes challenge: A retrospective. Int. J. Comput. Vision 111(1), 98–136 (2015)
53. Russakovsky, O., et al.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vision 115(3), 211–252 (2015)
54. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, Munich, Germany, pp. 234–241 (2015)
55. Honda, T., Sugimura, D., Hamamoto, T.: Multi-frame RGB/NIR imaging for low-light color image super-resolution. IEEE Trans. Comput. Imaging 6, 248–262 (2020)
56. Tian, C., et al.: Attention-guided CNN for image denoising. Neural Networks 124, 117–129 (2020)
57. Land, E.H., McCann, J.J.: Lightness and retinex theory. Josa 61(1), 1–11 (1971)
58. Jobson, D.J., Rahman, Z.U., Woodell, G.A.: Properties and performance of a center/surround retinex. IEEE Trans. Image Process. 6(3), 451–462 (1997)
59. Wang, S., et al.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Trans. Image Process. 22(9), 3538–48 (2013)
60. Fu, X., et al.: A fusion-based enhancing method for weakly illuminated images. Signal Process. 129, 82–96 (2016)
60. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. In: 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, Canada, pp. iii–709 (2004).

61. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a “completely blind” image quality analyzer. IEEE Signal Process Lett. 20(3), 209–212 (2013).

62. Huang, L., Zhao, X., Huang, K.: GOT-10k: A Large High-Diversity Benchmark for Generic Object Tracking in the Wild, in IEEE Transactions on Pattern Analysis and Machine Intelligence, Los Alamitos, CA, USA, doi: https://doi.org/10.1109/TPAMI.2019.2957464

How to cite this article: Li M, Zhou D, Nie R, Xie S, Liu Y. AMBCR: Low-light image enhancement via attention guided multi-branch construction and retinex theory. IET Image Process. 2021;15:2020–2038. https://doi.org/10.1049/ipr2.12173