Low-Light Image Enhancement Based on Retinex Theory and Attention Mechanism

Linlin Jiao[1,*], Fang Zhang[2]

{jiao120688@outlook.com[1], 1520404008@gmail.com[2]}

School of Artificial intelligence and Big Data, Henan University of Technology, Zhengzhou, China[1]
School of Computer Science&Technology, Huazhong University of Science and Technology, Wuhan, China[2]
*corresponding author

Abstract. With the popularity of mobile devices and surveillance cameras, image enhancement under low-light conditions has become an crucial research direction in computer vision. Although existing low-light enhancement techniques have made some progress, they still have limitations in processing complex scenes and maintaining image details. To handle these challenges, this paper introduces an approach on account of an improved Retinex-Net and attention mechanism. Our approach combines the retinex theory with Squeeze-and-Excitation Networks (SENet) and the total variance (tvloss) loss function to enhance the quality of low-light images. Comparative experiments were conducted on the LOL dataset, and the results of the experiments confirm that our proposed improved model provides significant improvements in peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and subjective visual performance compared to the original Retinex-Net model. Ablation experiments were executed to analyze the role of each module in the proposed method individually. The empirical evidence supports the functionality of each module within the algorithm and the advancement of the overall method.

Keywords: low-light, retinex, attention, tvloss.

1 Introduction

Vision is an important sense for human beings to explore and understand the external world, and images, as the main carrier of visual information, contain rich environmental details. However, in the real world, owing to variables including visual presentation strategies or insufficient background light, the images we acquire are often dark, which not only reduces the aesthetics of the images, but also affects the execution of the subsequent computer vision tasks, including object tracking, image categorization, and semantic segmentation, etc [1]. In order to reveal illegible details under low-light conditions and advance the operation of computer vision systems, it is significant to propose a solution for improving the overall quality of images in low-light scenarios.

The approaches in this area can primarily be divided into conventional techniques and those utilizing deep learning and neural networks. Among the traditional methods, low-light image enhancement techniques chiefly encompass histogram equalization based methods [2] and gamma correction based methods [3].M. Abdullah-Al-Wadud et al. proposed an intelligent

CONF-MLA 2024, November 21, Adana, Turkey

DOI 10.4108/eai.21-11-2024.2354585

contrast enhancement technique based on the traditional histogram equalization algorithm [4] in 2007.This technique is known as Dynamic Histogram Equalization, which can enhance image contrast without loss of image details by controlling the traditional histogram equalization, it can enhance the contrast of an image without loss of image details.Jeon Jong Ju et al. proposed a model for low-light image enhancement using gamma correction in a mixed color space [5], and also proposed a pixel-adaptive gamma determination algorithm to prevent under-enhancement or over-enhancement. The advantage of this method is that it does not require a training or refinement process and therefore is fast. In addition to this, methods based on retinex theory [6-8] decompose color images into reflectance and illuminance and use the principle of color constancy to recover low-quality pictures.

The progress in deep learning has established deep learning-oriented methods such as CNN and GAN as the new frontier in image enhancement. In 2017, Li Tao, Chuang Zhu et al. proposed the LLCNN [9] neural network based on CNN. The architecture design of LLCNN allows learning multi-scale feature maps and effectively avoids the problem of gradient vanishing during training. The CNN-based Zero-DCE neural network [10], proposed in 2020, is a novel reference-free method to improve the quality of low-light images without pairwise training data. MBLLEN [11], a deep learning method for low-light image and video enhancement, was initially proposed in a paper in 2018, and the method improves image quality by utilizing a multi-branch network structure to extract features at different levels. EnlightenGAN [12] is an unsupervised generative adversarial network whose network structure mainly consists of a U-Net generator with a self-attention mechanism and a pair of global-local discriminators. In addition, EnlightenGAN employs a self-feature preservation loss to guide the training process, allowing it to preserve texture and structure information. These neural network-based, deep learning methods can substantially improve the vividness and contrast of images with insufficient light by learning the deeper features of an image, and can even produce visually realistic images. In this context, Wei et al. from Peking University proposed the novel image processing network of Retinex-Net [13-15] in 2018, which, as an innovative framework combining retinex theory and deep convolutional neural networks, significantly improves the quality of images in low-brightness environments through its unique decomposition and augmentation network structure. However, the network is not designed with a unique structure in terms of feature extraction, which sometimes results in a less adaptive network that struggles to extract features adequately. Looking ahead, continued optimization and research will further enhance the performance of Retinex-Net and expand the boundaries of its applications in image enhancement.

To sum up, we put forward a method for enhancing images in dark settings, which leverages retinex theory and attention mechanism. By harnessing the interpretability of retinex theory and the feature extraction proficiency of attention mechanisms, the model can capture more salient features and attain more pronounced image enhancement.

Traditional low-light image enhancement techniques share the advantage of not requiring additional training data, which makes them efficient for real-time applications. However, they tend to be less efficient when processing large images due to increased computational demands. Models like MBLLEN and LLCNN excel at learning multi-scale features, enabling them to capture a rich tapestry of image details. Yet, their effectiveness is contingent upon the quality and diversity of the training data. Zero-DCE employs a lightweight DCE-Net architecture that is not only parameter-frugal but also computationally efficient, yet it may struggle with images

featuring uneven illumination. Retinex-Net stands out as an end-to-end trainable network, capable of directly extracting and learning useful features from data. However, it can amplify noise in the darker regions of the image post-reflection component extraction, potentially degrading image quality. Building upon Retinex-Net, the new model incorporates SENet and tvloss to further enhance its capabilities. SENet directs the network's focus towards prominent image features while downplaying the insignificant ones, thereby refining the enhancement of details and overall image quality. Meanwhile, tvloss promotes spatial smoothness in the output images, aiding in the suppression of noise and contributing to cleaner, more polished low-light imagery.

2 Method

The new model proposed by us takes Retinex-Net as the basic architecture, integrates SENet's attention mechanism into the convolutional layer and introduces tvloss during the training process. The flowchart is depicted below:

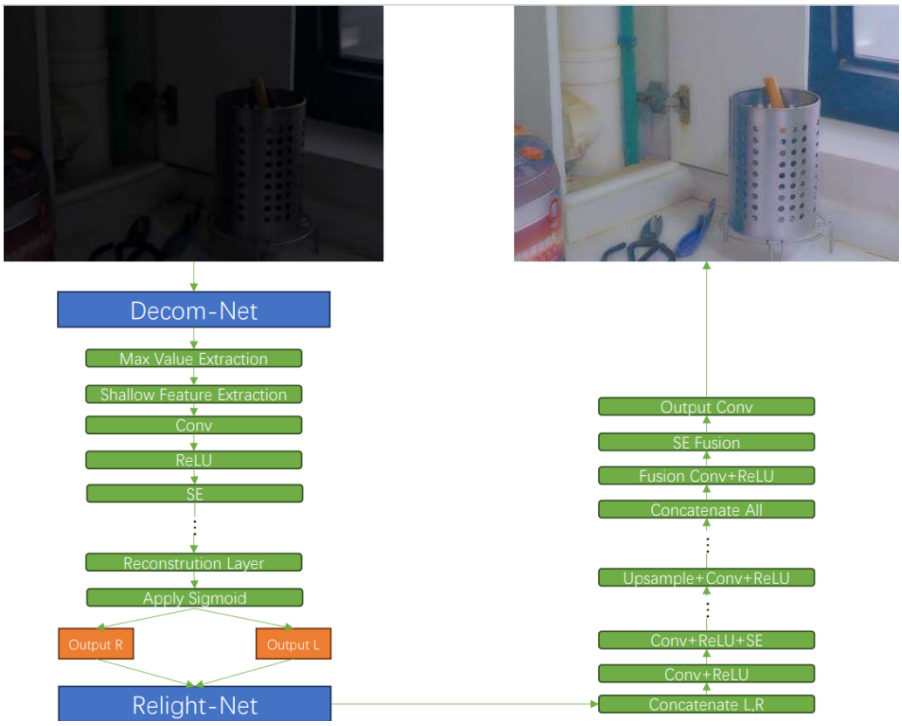

Fig. 1. Framework diagram of the new model

2.1 Retinex Theory

Retinex theory is a theory of visual perception originally proposed by American scientist Edwin H. Land in 1963. Under the retinex framework, the intensity of each pixel in an image is largely governed by the light intensity hitting the object, while the object's own reflective properties

define the intrinsic properties of the image. This implies that the image perceived is fundamentally the result of light reflected from an object under particular lighting conditions. From this perspective, an image can be decomposed into two parts: one part is the reflective properties of the object, and the other part is the lighting conditions of the environment. The formula for the expression mentioned above is:

$$S(x,y) = R(x,y) \times I(x,y) \tag{1}$$

Where S(x, y) represents the color image at a particular location (x, y), R(x, y) represents the reflected component at the same location and I(x, y) represents the illuminated component which are combined together by pixel-by-pixel multiplication.

Reflectance represents a stable property of an object that remains the same regardless of lighting variations. Lighting conditions, on the other hand, determine the degree of lightness and darkness on various parts of an object's surface. In poorly lit images, objects can often be seen to exhibit darker tones and uneven lighting effects.

For this purpose, the process of enhancing dark light pictures by retinex theory is mainly as follows:

(1) Decompose the input low-light image S Decompose it into a light image I and reflection image R

(2) The reflection image R After processing such as denoising and detail enhancement, the light image I is corrected to improve the visual clarity and quality of the image to obtain the adjusted reflection image R' and the illumination image I' The adjusted reflection image and illumination image are obtained:

$$R' = Net(R), I' = Net(I) \tag{2}$$

Where Net stands for the corresponding network process of luminance adjustment and denoising.

(3) The final enhanced image is obtained by fusing the adjusted light image with the recovered reflection image S' :

$$S' = merge(I', R') \tag{3}$$

2.2 Attention mechanisms

One of the most significant concepts in deep learning is the attention mechanism, which is based on the human biological system. Attention mechanisms can help us filter out irrelevant or useless information and focus on important stimuli, tasks or goals. The attention mechanism used in this paper is Squeeze-and-Excitation Networks abbreviated as SENet [16], which is a new network structure proposed by Momenta and Jie Hu et al. The working principle of SENet is to automatically obtain the importance of each feature channel by learning, and in light of this importance to foster the significant features and oppress the insignificant ones for the current task. SENet network structure has two key operations: Squeeze and Excitation.

(1) Squeeze: this step compresses the feature map into a single channel descriptor by Global Average Pooling, which compresses the spatial dimensions of each channel of the feature map U to obtain a global feature vector z of length C. The formula is as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} u_c\,(i,j) \tag{4}$$

where *uc* is the *cth* channel in the feature map *U*, *W* and *H* are the width and height of the feature map, respectively, and *zc* is the *cth* element in the global feature vector.

With the above operation, the information from each channel is aggregated to form a real number that can characterize the global feature response. This descriptor encompasses the information from the network's expansive receptive field, making the global receptive field available even to layers close to the input layer.

(2) Excitation: the excitation operation is implemented through two fully connected (FC) layers with the aim of learning the importance weights of each channel. First, z is reduced from C-dimension to C/r-dimension (where r is the scaling parameter) through the first FC layer, then the function is activated through ReLU, and then the dimension is upgraded from C/r back to C-dimension through the second FC layer, and finally the weight s is obtained through the sigmoid function. the formula is as follows:

$$s_c = \sigma(W_2 \delta(W_1 z)) \tag{5}$$

where W1 and W2 are the weights of the fully connected layer, δ is the ReLU activation function, σ is the sigmoid activation function, and *sc* is the *cth* element of the output weight vector. Finally, each channel of the original feature map is multiplied by its corresponding weight *sc* to obtain the feature map after SE block processing. The formula is as follows:

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \tag{6}$$

where *xc* is the *cth* channel of the output feature map after SE block processing and *uc* is the *cth* channel of the input feature map. With the Excitation module, SENet is able to explicitly model the interrelationships between the feature channels instead of implicitly capturing these relationships through a convolutional neural network, as in the case of traditional convolutional neural networks. filter to capture these relationships.

3 Experiment

3.1 Experimental environment

The tests were processed on a laptop containing an NVIDIA GeForce RTX 3060 Laptop GPU, and all experiments were programmed in a Python 3.9.19 environment, using the Pytorch architecture 1.12.0 for the deep learning tasks, and Pycharm Community Edition 2020.1.3 x64 developed as an integrated development environment.

3.2 Dataset and training details

The training set for this experiment consists of 485 pairs of real scene image pairs from the LOL dataset [13] and 1000 pairs of synthesized images, and the test set consists of the remaining 15 pairs of images from the LOL dataset.

Fig. 2. Two pairs of datasets in the LOL dataset

Fig. 3. Two pairs of datasets in synthesize low-light images

In the training process, the network is trained using stochastic gradient descent (SGD) with an initial learning rate set to 0.001, adjusting the learning rate to one-tenth of the initial learning rate from the 21st epoch onwards, with the batch_size size set to 16 and the patch_size size set to 96 × 96.

3.3 Evaluation indicators

The evaluation metrics used in this experiment are PSNR and SSIM.The formulae for PSNR and SSIM are as follows:

$$PSNR = 20 \cdot log_{10}(\frac{MAX_1}{\sqrt{MSE}}) \tag{7}$$

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{8}$$

where MAXI is the image's top pixel value. μx and μy are the mean of the image x and y, respectively, σx and σy are the standard deviation of the image x and y, respectively, and σxy is the covariance of the image x and y, respectively. C1, C2 and C3 are constants used to maintain stability. Among them, PSNR (Peak Signal-to-Noise Ratio) is a pixel-based image quality assessment index, which measures the image quality by comparing the mean square error (MSE) between the original image and the distorted image, and the higher the value of PSNR, the better the quality of the image, and the smaller the distortion.SSIM (Structural Similarity Index (SSIM) is a kind of image quality assessment index that is more in line with the visual characteristics of the human eye, which evaluates image similarity by assessing the luminance, contrast, and structural details of two images, with scores ranging from 0 to 1, where a score closer to 1 indicates greater image similarity.

3.4 Comparative experiments

Qualitative analysis. The difference in image enhancement between the improved method and the original Retinex-Net method is first demonstrated through visual comparison. Figure 3.3 shows a set of enhanced image comparisons.

Fig. 4. Comparison of the enhanced images of the two methods, with the original method on the left and our method on the right

As shown in the figure, our improved method is notably better at enhancing image details and contrast, and is more capable of restoring the image's natural color and detail level.

Quantitative analysis. Table 3.1 demonstrates the comparison of our method with the original Retinex-Net method on both PSNR and SSIM metrics.

Table 1. Comparison of PSNR and SSIM values

Method	PSNR(dB)	SSIM
Retinex-Net	17.5577091000428	0.644808102405132
Our method	18.3186832929728	0.678646289901253

The table's details confirm that the improved algorithm outperforms the original Retinex-Net method in both PSNR and SSIM metrics, which indicates a significant improvement in the quality of image enhancement.

3.5 Ablation experiments

The experimental results were evaluated under three conditions: using the original Retinex-Net model, incorporating SENet, and incorporating both SENet and tvloss. The evaluation measures are reported in Table 3.2 herein:

Table 2. Comparison of PSNR and SSIM values for three experiments

Method	PSNR(dB)	SSIM
Retinex-Net	17.5577091000428	0.644808102405132
Retinex-Net+SENet	18.0428951203385	0.689933900276609
Retinex-Net+SENet+tvloss	18.3186832929728	0.678646289901253

The data in the table clearly indicates a significant improvement in PSNR and SSIM values after incorporating SENet, with the scores rising to 18.043and 0.690, respectively. However, upon the addition of tvloss, although the PSNR value further increased to 18.319, the SSIM value experienced a slight decrease, dropping to 0.679. From the comparison images, it can be seen that the new model enhances the overall image quality and improves the details, and effectively suppresses the noise.

4 Conclusion

We present an innovative enhancement of the Retinex-Net model that integrates Squeeze-and-Excitation Networks (SENet) and total variance (tvloss) loss function for the image enhancement problem under low-light conditions. Our approach uses the Retinex-Net network as the basic network architecture, inserts the SENet module on top of it to increase the model sensing field and fully extract features, and introduces the tvLoss loss function as well to optimize the model's training. Experimental results on the LOL dataset show that our method outperforms existing low-light image enhancement techniques in terms of objective metrics such as PSNR SSIM and subjective visual quality. These results demonstrate the ability of SENet to enhance feature representation and the important role of tvloss in reducing artifacts

and noise. Subsequent work should focus on studying more complex network structures to improve the effectiveness and generality of the model.

References

[1] Guo, J., Ma, J., García-Fernández, G.F., Zhang, Y., & Liang, H. (2023). A survey on image enhancement for low-light images. Heliyon, 9 (4).

[2] Yuan, Q., Dai, S. (2024). Adaptive histogram equalization with visual perception consistency. Information Sciences, 668.

[3] Xu, G., Su, J., Pan, H., Zhang, Z., Gong, H. (2009). An image enhancement method based on gamma correction. 2009 Second International Symposium on Computational Intelligence and Design (pp. 60-63).

[4] Abdullah-AI-Wadud, M., Kabir, M.H., Dewan, M.A.A., Chae, O. (2007). A dynamic histogram equalization for image contrast enhancement. IEEE Transactions on Consumer Electronics (pp. 593-600).

[5] Jeon, J., Park, J., Eom, I. (2024). Low-light image enhancement using gamma correction prior in mixed color spaces. Pattern Recognition, 146.

[6] Gu, Z., Li, F., Fang, F., Zhang, G. (2020). A novel retinex-based fractional-order variational model for images with severely low-light. IEEE Transactions on Image Processing (pp. 3239-3253).

[7] Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y. (2023). Retinexformer: One-stage retinex-based transformer for low-light image enhancement. Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 12504-12513).

[8] Hao, S., Han, X., Guo, Y., Xu, X., Wang, M. (2020). Low-light image enhancement with semi-decoupled decomposition. IEEE Transactions on Multimedia (pp. 3025-3038).

[9] Tao, L., Zhu, C., Xiang, G., Li, Y., Jia, H., Xie, X. (2017). LLCNN:A convolutional neural network for low-light image enhancement. 2017 IEEE Visual Communications and Image Processing (pp. 1-4).

[10] Guo, C., Li, C., Guo, J., Chen, C., Hou, J., Kwong, S., Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1780-1789).

[11] Lv, F., Lu, F., Wu, J., Lim, C. (2018). MBLLEN: Low-light image/video enhancement using cnns. British Machine Vision Conference, 220(1).

[12] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z. (2021). Enlightengan: Deep light enhancement without paired supervision. IEEE Transactions on Image Processing (pp. 2340-2349).

[13] Wei, C., Wang, W., Yang, W., Liu, J. (2018). Deep retinex decomposition for low-light enhancement. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[14] Hai, J., Hao, Y., Zou, F., Lin, F., Han, S. (2023). Advanced retinexnet: a fully convolutional network for low-light image enhancement. Signal Processing: Image Communication,112.

[15] Anoop, P, P., Deivanathan, R. (2024). Advancements in low light image enhancement techniques and recent applications. Journal of Visual Communication and Image Representation,103.

[16] Hu, J., Shen, L., Sun, G. (2018). Squeeze-and-excitation networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7132-7141).