
LOW-LIGHT IMAGE ENHANCEMENT IN THE FREQUENCY DOMAIN

Hao Chen
Sun Yat-sen University
chenh366@mail2.sysu.edu.cn

Zhi Jin
Sun Yat-sen University
jinzh26@mail.sysu.edu.cn

ABSTRACT

Decreased visibility, intensive noise, and biased color are the common problems existing in low-light images. These visual disturbances further reduce the performance of high-level vision tasks, such as object detection, and tracking. To address this issue, some image enhancement methods have been proposed to increase the image contrast. However, most of them are implemented only in the spatial domain, which can be severely influenced by noise signals while enhancing. Hence, in this work, we propose a novel residual recurrent multi-wavelet convolutional neural network (**R2-MWCNN**) learned in the frequency domain that can simultaneously increase the image contrast and reduce noise signals well. This end-to-end trainable network utilizes a multi-level discrete wavelet transform to divide input feature maps into distinct frequencies, resulting in a better denoise impact. A channel-wise loss function is proposed to correct the color distortion for more realistic results. Extensive experiments demonstrate that our proposed R2-MWCNN outperforms the state-of-the-art methods quantitatively and qualitatively.

1 Introduction

Low-light image enhancement is critical for many vision tasks, including recognition [1] and nighttime autonomous driving [2]. However, low-light scenarios introduce a succession of visual degradations, e.g. diminished visibility, intensive noise and color distortion in captured images. These images suffer from information loss and are likely to undermine the performance of algorithms that are primarily designed for high-visibility conditions. Therefore, to make the buried information visible, as shown in Figure 1, low-light image enhancement is expected. However, it is non-trivial to enhance low-light images, as the noise and color distortion are easily be amplified unconsciously.

Many researchers have focused on addressing these issues over the last few decades, and many effective algorithms have been proposed. Traditional low-light image enhancement algorithms can be broadly classed as Histogram Equalization-related [3] and Retinex theory-related [4] methods. Histogram Equalization (HE) and its derivatives [5, 6], which mainly focus on enhancing the contrast of images at the pixel level, can unintendedly amplify blind noise while restoring illumination. The Retinex theory-based algorithms [7, 8] decompose image into a constant reflectance layer and a variant illumination layer, with the purpose of restoring the illumination layer of the image independently. In practice, however, reflectance and illumination layers are roughly estimated, which can cause visible color distortion and further decrease visual quality.

The recent blossoming of deep learning has inspired the development of low-light enhancement. Some approaches [12, 8] are built on the Retinex theory, while some others design CNNs [13, 14] and GANs [15] to enhance low-light images directly. However, the shared problem is that these methods are mainly designed in the pixel domain, which suits human perception rather than machine vision.

Precisely, for computer vision, the random noise carried by low-light images is hardly to be separated in the pixel domain. Most of the existing denoising algorithms perform an individual filtering process or noise estimation elimination, and can efficiently reduce the noise levels. However, the images are either blurred or over-smoothed, resulting in unsatisfactory details and visual quality [16, 17]. Therefore, although recalling the visibility in the pixel domain is the most intuitive and rather effective way, their capabilities for noise suppression and detail restoration are limited.

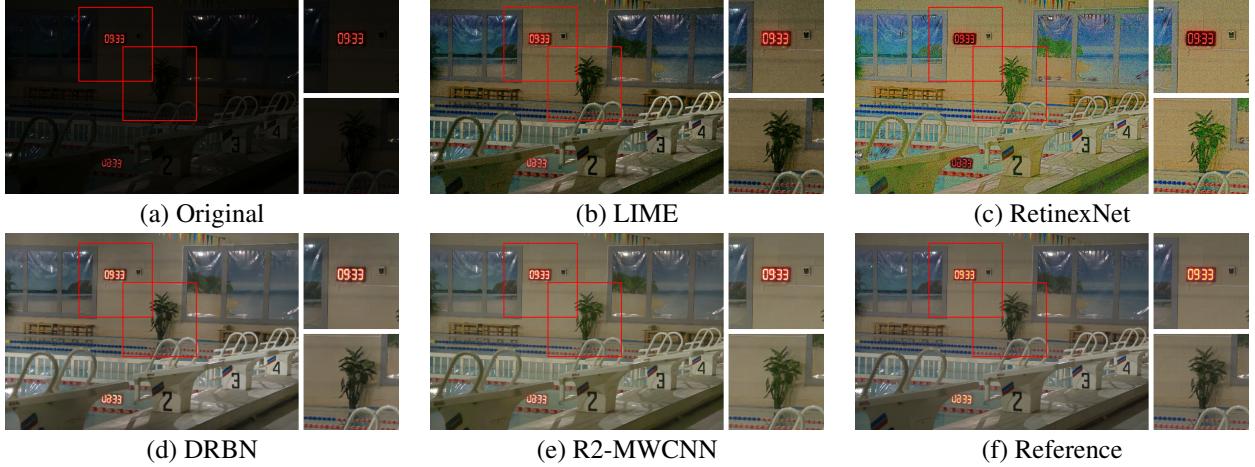


Figure 1: The enhanced results in *LOL-Real* dataset [9] of different methods (PSNR/SSIM). (a) Low-light input. (b) LIME [10] (19.2096/0.5778). (c) RetinexNet [8] (17.0594/0.4011). (d) DRBN [11] (25.8050/0.9273). (e) R2-MWCNN (27.8149/0.9378). (f) Normal light reference. The proposed R2-MWCNN well restores high quality images from real low-light inputs in the *LOL-Real* dataset.

Considering the separability between noise and structural information in the frequency domain [17], in this paper, we propose a frequency-based architecture to address the problems of enhancing low-light images in the pixel domain. The core of our architecture is a novel residual recurrent multi-wavelet convolutional neural network (R2-MWCNN). The Discrete Wavelet Transform (DWT) and Inverse Discrete Wavelet Transform (IDWT) [18] are introduced to replace commonly used pooling and up-sampling operations and exploit multi-scale frequency-based features for denoising and detail restoration. We also propose the Multi-level Short-cut Connections (MSC) module to select and transmit low-level semantic information adaptively for detail recovery. Moreover, since enhanced low-light images usually suffer from severe color distortion, a channel-wise loss function is proposed to aid the retrieval of pleasing visual quality. Overall, our contributions are as follows:

- We propose a novel end-to-end frequency-based architecture for enhancing low-light images. It integrates illumination enhancement, color restoration and denoising, and can improve both the objective and subjective image quality.
- We propose an effective low-light image enhancement network with the proposed MSC modules to adaptively select the high-/low-level semantic features for satisfactory detail restoration. DWT and IDWT are introduced to exploit multi-scale frequency-based features, which serves as an appropriate basis for separating noise from the structural information.
- A novel channel-wise loss function is designed to handle the pixel-intensity distribution of each color channel in order to overcome color distortion and improve the visual quality. Extensive experiments show that our framework outperforms the existing state-of-the-art methods both quantitatively and qualitatively.

2 Related Work

Low-light enhancement techniques can be classified into classic or deep learning-based algorithms according to whether a data-driven deep network is adopted.

Classic enhancement algorithms. Histogram equalization and its derivatives are one of the classic algorithm categories. CLAHE [19] attempts to alleviate noise amplification by limiting the range of contrast. BPDHE [20] aims to maintain brightness while dynamically correcting contrast. LDR [21] proposes investigating the layered difference representation of 2-D histograms in order to enhance contrast by emphasizing gray-level differences between adjacent pixels. DHECE [22] combines differential gray-levels histogram equalization of saturation and intensity to generate appropriate high-contrast results.

Another traditional low-light enhancement category is built based on the Retinex theory, which assumes an image consists of a constant reflectance layer and a variant illumination layer. Hence, if these two layers can be successfully separated and enhanced properly, the low-light image can be enhanced well by combining them. SSR [23] and MSR [24] are typical Retinex theory-based enhancement algorithms that attempt to rectify the illumination map for brightness

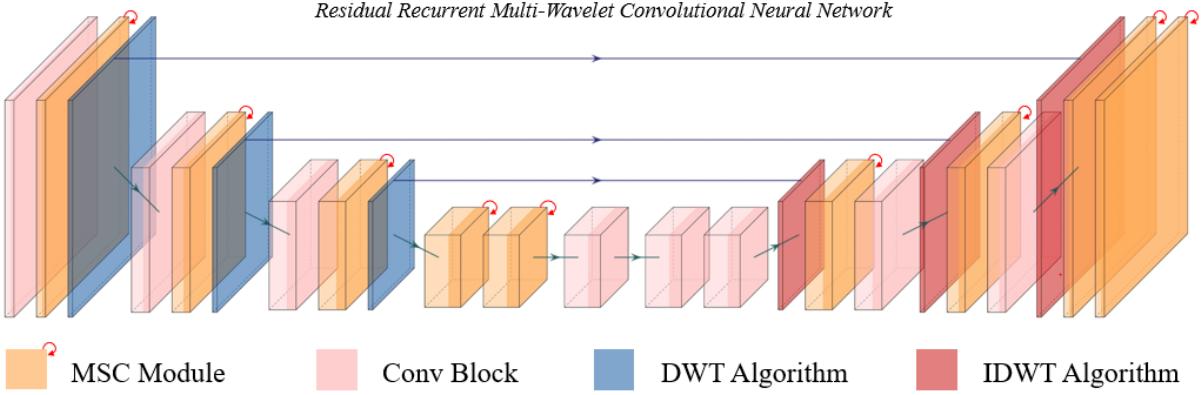


Figure 2: The end-to-end architecture of the Residual Recurrent Multi-Wavelet Convolutional Neural Network (R2-MWCNN) with U-Net [36], wavelet transform and Multi-level Short-cut Connections module (red swirling arrow).

enhancement. AMSR [25] proposes a weighting strategy for each MSR output. MSRCR [26] tries to enhance reflectance estimation with a color restoration factor. RRM [27] designs an optimization function to constrain noise in illumination estimates. SRIE [28] employs a weighted vibrational model to estimate reflectance and illumination. LIME [10] refines the maximum-approximated illumination and recomposes the enhancement result in accordance with Retinex theory.

However, these methods heavily rely on manually designed parameters that suit different low-light images. Hence, they are not flexible and easily amplify noise and color distortion, causing decreased visual quality.

Deep learning-based methods. Recently, deep learning has stimulated tremendous progress in low-light image enhancement. For example, LLNet [29] is the first deep learning trial in low-light image enhancement, proposing an auto-encoder for contrast enhancement and denoising. MSR-net [30] is the earliest implementation of the Retinex theory-based supervised CNN. EnlightenGAN [31] is a successful trial of Generative Adversarial Nets (GANs) that can adaptively learn from unpaired low/normal-light images. Retinex-Net [8] is designed to integrate deep learning with Retinex theory, which can decompose the image to enhance the illumination and reflectance, respectively. DeepUPE [32] presents a neural network that enhances under-exposure images by learning complicated photography adjustments. SICE [33] includes an end-to-end CNN to adapt the contrast of multi-exposure input images. MBLLEN [34] adopts three modules within a single network, improving the output of each extraction layer and combining them together to create the final enhanced image. DRBN [11] proposes a semi-supervised learning framework to enhance low-light images from coarse to fine.

Although these methods have demonstrated outstanding low-light enhancement performance, due to the limitations of the pixel domain, blind noise can not be effectively eliminated from the images. Blurring and color distortion, which are caused by spatial denoising, continue to obstruct practical application. To overcome this shared limitation, R2-MWCNN employs an invertible wavelet transform [35] to perform denoising in the frequency domain so that to effectively recover sharp structures from degraded observation. In real-world scenes, the image information has a certain structure, but the noise is random without a specific structural property. Hence, the introduction of the wavelet transform allows the network to separate the image structural information and the noise, which aids in preserving pleasing details during denoising. Moreover, a novel channel-wise loss is proposed to correct the common color distortion problem in low-light enhancement.

3 Methodology

3.1 R2-MWCNN Network

Our network comes from two inspirations. First, it is easier to remove noise from image structural information in the frequency domain and to estimate the global illumination from the low-frequency features [37]. Second, it is known that as the network depth increases, the model can extract higher-level semantic information from the image. Nevertheless, with the common employment of down-sampling operations, the original low-level structural information may be lost as the network deepens. Hence, as shown in the Figure 2, the proposed R2-MWCNN is built with the U-Net structure [36], wavelet transform and several MSC modules. The MSC module is employed by replacing normal

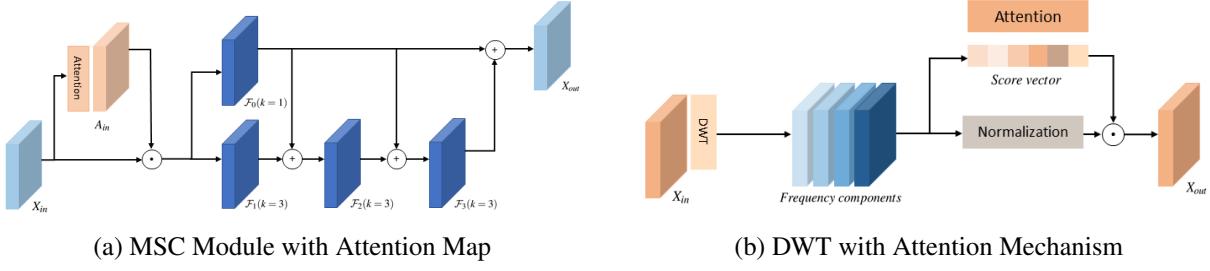


Figure 3: The MSC module with attention map utilizes several shortcut connections and multiple add operations to maintain low-level original graphic information. The DWT with attention mechanism performs space-to-frequency transform, which enables the model to separate noisy signals from structural information.

covolutional blocks, which can transmit low-level graphic information into deeper network layers without requiring additional computational cost. Our R2-MWCNN comprises a contracting subnetwork and an expanding subnetwork. In the contracting subnetwork, the invertible DWT is introduced to replace the frequently used pooling operation and perform space-to-frequency transform. In the expansion subnetwork, IDWT is presented to restore all the information that is kept by DWT to high resolution in the pixel domain.

MSC module. Inspired by recurrent convolutional layer [38], we propose the MSC module, as shown in Figure 3 (a). The goal of the MSC module is to preserve and transmit low-level features into subsequent layers without requiring additional processing resources, so as to restore satisfactory details. Given the input features $X_{in} \in \mathcal{R}^{H \times W \times C}$ and convolutional operation $\mathcal{F}(\cdot)$, the MSC module first calculates the pixel-wise contrast-aware attention map C_a [39] to select features by:

$$\begin{aligned} C_a &= \text{sigmoid}(\mathcal{F}_{map}(X_{in})), \\ X_a &= C_a \cdot A_{in}. \end{aligned} \quad (1)$$

Next, considering the information loss as the network deepens, several shortcut connections are conducted between the input layer and the subsequent convolutional layers in the MSC module to adaptively select low-level graphic information. The shortcut connection [40] performs an add operation between the identity mapping and the output of previously stacked features, which can be calculated by $\mathcal{F}(x_1) + x_2$. Considering that the output of the l^{th} layer in the MSC module is X_{l+1} , which can be calculated accordingly:

$$X_{l+1} = \mathcal{F}_l(X_l + \mathcal{F}_0(X_a)), \quad (2)$$

where $\mathcal{F}_0(\cdot)$ represents 1×1 convolution of input X_a to match the output dimensions.

Wavelet Transform. To guarantee that all the information can be well-preserved, in the R2-MWCNN, we utilize DWT and IDWT to replace the commonly-used downsampling and upsampling. The DWT can be viewed as a top-down collection of several filters that can theoretically extract both frequency and location features, shown in Figure 3 (b), while IWT is employed to substitute the transpose convolutional layer to integrate low resolution features maps into recovered high resolution ones.

In practice, given the trade-off between fidelity gain and computational cost, a two-layer DWT with four sub-band outputs (X_{LL}, X_{LH}, X_{HL} and X_{HH}) is adopted [18], where the subscript index is an ordered representation of filter processing. For example, assuming the input sample is $X_{in} \in \mathcal{R}^{H \times W \times C}$, X_{LL} is the output of X that passes through two low-pass filters, which can be calculated as below:

$$X_{LL} = X_{in} \otimes f_{LL}, \quad (3)$$

where f_{LL} can be defined as:

$$f_{LL} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \quad (4)$$

Analogously, the rest of the filter definitions are shown as follows:

$$f_{LH} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, \quad f_{HL} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, \quad f_{HH} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}. \quad (5)$$

As can be seen, the pixel-intensity of the four sub-band outputs is substantially distinctive. Hence, we normalize the value range to $[0, 1]$ to ease computational burden. Finally, the attention mechanism [41] is introduced to adaptively scale the feature maps from different sub-bands in a channel-wise manner.

3.2 Loss Function

The Mean Square Error (MSE), also known as L2 loss, is a commonly used error measurement for low-light image enhancement. Nevertheless, L2 loss has been proven to be insufficient for recovering favourable details and can even induce blur at the object edges [42]. In order to finely recover the visual details of the image, we use multiple losses to measure differences in various aspects of the enhanced image and the corresponding ground truth. The whole loss function for training consists of four parts:

$$\text{Loss} = L_{pixel} + L_{global} + L_{edge} + L_{channel} \quad (6)$$

Pixel-wise Loss. The pixel-wise loss, indicated as L_{pixel} , is a variant of the Smooth L1 loss. The wider gap between blind noise and structural information in Smooth L1 loss compared to L2 loss, makes the network more focused on denoising. Experiments reveal that our network performs well in dark regions yet over-enhances bright regions. After conducting several trials, the images are divided into bright and dark region with a 3:7 area ratio. Different weights w_1 and w_2 are assigned to provide a better approximation of the bright region.

$$\begin{aligned} L_{pixel} &= w_1 \cdot \frac{1}{n_L m_L} \sum_{i=1}^{n_L} \sum_{j=1}^{m_L} \text{SmoothL1}(E_L(i, j), G_L(i, j)) \\ &\quad + w_2 \cdot \frac{1}{n_H m_H} \sum_{i=1}^{n_H} \sum_{j=1}^{m_H} \text{SmoothL1}(E_H(i, j), G_H(i, j)), \end{aligned} \quad (7)$$

where E_H and G_H represent the bright region of the enhanced image and the ground truth reference, and E_L and G_L are the rest of the image.

Global-wise Loss. The global-wise loss is a combination of structure loss and perceptual loss [43]. Structural loss is calculated using Structure Similarity (SSIM) [44], which is designed to assess the global variations in contrast, brightness, and structure between the enhanced image and the normal image from a global perspective. The perceptual loss is offered to primarily concerned with high-level context via the employment of feature-extraction VGG network [45]. If the enhanced results and the original images are close, the VGG extractor outputs should be similar as well. Our global-wise loss is calculated as follows:

$$L_{global} = w_3 \cdot \frac{1}{C_{i,j} H_{i,j} W_{i,j}} \|\phi_{i,j}(E) - \phi_{i,j}(G)\|_2^2 - w_4 \cdot \frac{(2\mu_E\mu_G + C_1)(2\sigma_E\sigma_G + C_2)}{(\mu_E^2 + \mu_G^2 + C_1)(\sigma_E^2 + \sigma_G^2 + C_2)}, \quad (8)$$

where E and G represent enhanced image and ground truth, respectively, and $\phi_{i,j}(\cdot)$ refers to the feature extractor of the i_{th} block j_{th} layer in the VGG network. $C_{i,j}$, $W_{i,j}$ and $H_{i,j}$ describe the 3-dimensions of feature maps extracted from the VGG-19 network. Towards the structure loss, μ_x and σ_x are average and variance, respectively, while σ_{xy} represents covariance between images. C is a small constant to prevent division by zero.

Edge Loss. Normally, denoising operations can introduce severe blurring. Words and object edges could be blurred after enhancement, which reduces the practicality of the low-light enhancement algorithms. With respect to that visual deterioration, the edge loss is utilized to evaluate the generated edges. In practice, the Sobel gradient is utilized as the edge information in the images. Two Sobel calculators, S_x and S_y , with x-axis and y-axis orientations, respectively, are used to estimate the gradient between two adjacent pixels. Thus, the edge loss between E and G can be calculated as follows:

$$L_{edge} = \|S_x \otimes E - S_x \otimes G\|_1 + \|S_y \otimes E - S_y \otimes G\|_1. \quad (9)$$

Channel-wise Loss. The previous methods generally use element-wise differences like L2 loss to measure the color fidelity. However, these methods often have color channel prejudice, resulting in generating color-biased images. The proposed channel-wise loss can make the network learn the pleasing global color distribution from bright images, which can guide low-light image enhancement. Therefore, satisfactory colors can be achieved. Firstly, Gaussian blur is used to eliminate image textures with kernel K_G :

$$K_G(k, l) = A \cdot \exp\left(-\frac{(k - \mu_x)^2}{2\sigma_x^2} - \frac{(l - \mu_y)^2}{2\sigma_y^2}\right), \quad (10)$$

where A is defined as 0.2, and Gaussian blurred image of E is referred to as E_b :

$$E_b(i, j) = \sum_i^{W-k} \sum_j^{H-l} E(i+k, j+l) \cdot K_G(k, l), \quad (11)$$

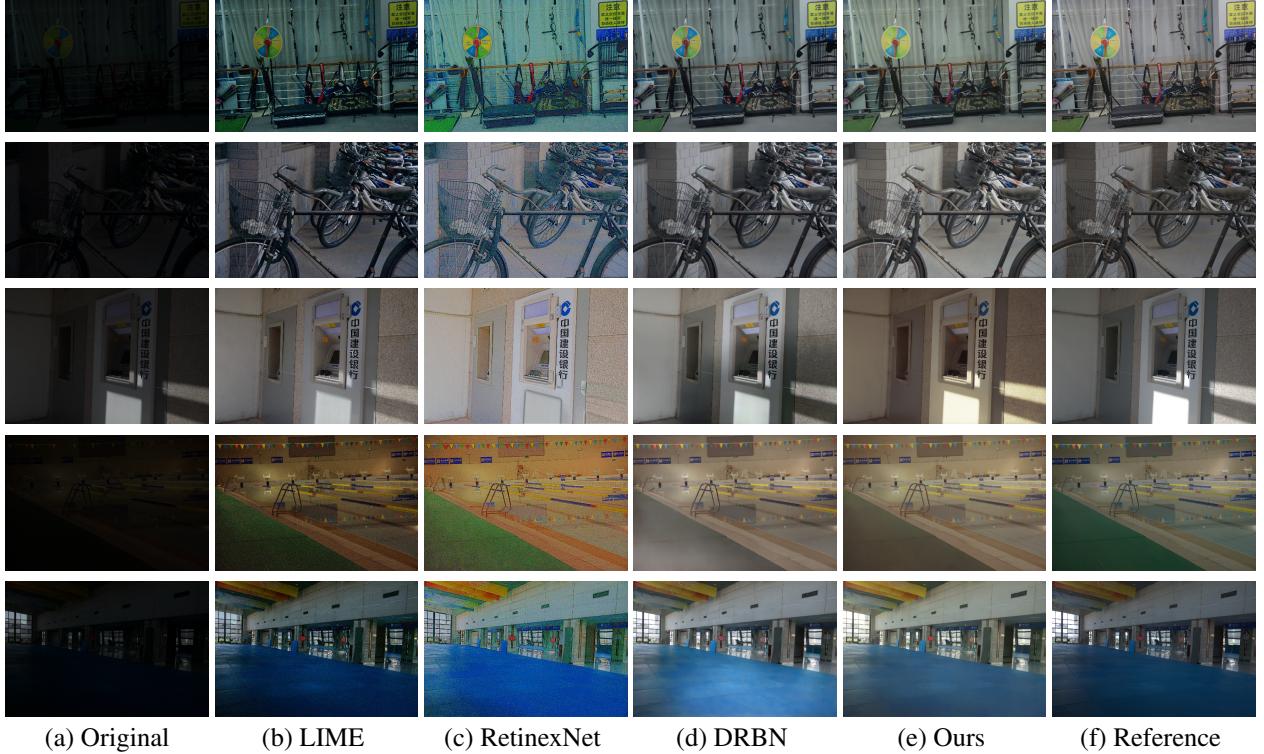


Figure 4: The visual results of state-of-the-art methods and ours on real low-light images from LOL-Real test dataset (top-to-bottom).

the operation towards G is the same. Hence, the channel-wise loss can be calculated by:

$$L_{channel} = \sum_h^C \left\| \sum_i^W \sum_j^H E_b(i, j, h) - \sum_i^W \sum_j^H G_b(i, j, h) \right\|_1. \quad (12)$$

The channel-wise loss can suppress biased color, which is typically produced by over or under enhancement of one or more of the RGB channels. Therefore, the network that is guided by this channel-wise loss can retrieve more real colors.

4 Experiment

4.1 Implementation Setting

We have implemented the proposed model in the Tensorflow framework [46]. The R2-MWCNN is optimized by the ADAM optimizer for 300 epoches with randomly initialized parameters. The initial learning rate and batch size for training are set to $2e^{-4}$ and 2. In order to reduce the computational cost, the input image values are scaled to $[0, 1]$. The learning rate is automatically decayed by the Tensorflow ReduceLROnPlateau with a patience of 10 and a decrease factor of 0.2. Rotation and flip based data augmentation have been adopted. The entire training process takes around 1 day to converge with the NVIDIA Tesla P40.

4.2 Comparison with Existing Methods

We conduct extensive experiments using real instances from the LOL-Real dataset [9]. Results are compared between R2-MWCNN and the other different state-of-the-art methods, including both classic and deep learning-based. Representative results are visually shown in Figure 4. In general, most previous methods implemented in the pixel domain have limited denoising and structure recovery capacities. By checking the details, BPDHE, LIME perform well in stretching contrast. Nevertheless, blind noise is heavily amplified, resulting in severe visual degradation. The great performance of DRBN, CRM and SICE in the objective evaluation shows their ability to restore global illumination and visibility. However, structural information and image details are not promising. Comparatively, R2-MWCNN can detach both the additive

and non-additive noise from the image structural information in the frequency domain effectively. In this way, our method achieves good detail fidelity and perceptual visual quality, as well as satisfactory global illumination.

Table 1: Quantitative results in *LOL-Real* dataset [9]. (Red: The best, Blue: The second)

Metric	BIMEF [47]	BPDHE [20]	CRM [48]	RetinexNet [8]	EFF [49]	LDR
PSNR	17.85	13.84	19.65	16.10	17.85	17.17
SSIM	0.6353	0.3990	0.6344	0.5689	0.6353	0.5697
FID	1.5623	2.7219	0.6480	1.8457	1.5623	2.7266
Metric	JED [50]	LIME [10]	MF [51]	RRM [27]	E-GAN [31]	DONG
PSNR	17.23	15.24	18.73	17.35	18.64	17.26
SSIM	0.7256	0.4181	0.5050	0.7251	0.6786	0.4755
FID	1.8785	11.908	0.9072	1.7040	0.7902	1.1610
Metric	ASMR	MSR [26]	DeepUPE [32]	SICE [33]	DRBN [11]	R2-MWCNN
PSNR	15.37	11.67	13.17	19.40	20.11	23.23
SSIM	0.4310	0.4064	0.4509	0.6842	0.8495	0.8648
FID	2.3692	7.8184	1.1641	0.6233	0.5039	0.2716

Table 1 contains a quantitative comparison of various methods. We adopt three objective image quality representatives as performance criteria: Structure Similarity (SSIM), Peak Signal-to-Noise Ratio (PSNR) and Frechet Inception Distance (FID) [52]. PSNR is utilized to measure the detail fidelity between the enhanced image and the normal light image. SSIM is more concerned with structural information and global illumination recovery. Consider that the perceptual visual quality might not be well captured by the PSNR and SSIM. We utilize FID score to measure the similarity between the enhanced image and the groundtruth image from the high-level context. The lower FID value indicates a greater quantity of similarity. Our method (Red) outperforms all the other methods in PSNR, SSIM and FID by a large margin. More concretely, R2-MWCNN considerably surpasses the second best method (Blue) in all three criteria, which means our method is equipped with better denoising, structural information restoration and can achieve very good visual fidelity.

4.3 Ablation Study

In this section, we analyse the effect of each technical design to support our motivation. Several ablation studies have been conducted to demonstrate that our method is promising to achieve satisfactory enhanced results. Table 2 shows the objective results (PSNR/SSIM) of the MWCNN and the R2-MWCNN with different loss function combinations.

Table 2: Ablation study of proposed methods on R2-MWCNN and MWCNN [18]

Metric	L_{pixel}	L_{global}	L_{edge}	$L_{channel}$	PSNR(dB)	SSIM
MWCNN [4]	✓	✓	✗	✗	19.8476	0.8390
	✓	✓	✓	✗	20.9474	0.8494
	✓	✓	✗	✓	21.4889	0.8478
	✓	✓	✓	✓	22.5147	0.8529
R2-MWCNN	✓	✓	✗	✗	20.9446	0.8517
	✓	✓	✓	✗	21.8730	0.8544
	✓	✓	✗	✓	22.3621	0.8512
	✓	✓	✓	✓	23.2260	0.8648

In each experiment, we turn off one or more of the loss functions to study their role in the enhancement. The same combination comparison between R2-MWCNN and MWCNN reveals that R2-MWCNN performs better in denoising and detail fidelity. In each block, the first row with previous loss functions serves as the baseline for comparison. The second and third rows serve as separate verifications of the proposed methods. The last row results show that the network with our loss functions achieves the best low-light enhancement performance.

5 Conclusion

In this paper, we design a frequency-based framework for low-light image enhancement. DWT is introduced to make the network learn to remove noise from the structural information. The proposed MSC module can adaptively select

low-/high-level semantic information for satisfactory detail reconstruction. We have also presented a channel-wise loss for color correction in low-light enhancement. Finally, we have conducted extensive experiments to verify the effectiveness of our method against state-of-the-art methods.

References

- [1] Alex Robert Wade and Frederick W Fitzke. A fast, robust pattern recognition system for low light level image registration and its application to retinal imaging. *Optics Express*, 3(5):190–197, 1998.
- [2] Hazem Rashed, Mohamed Ramzy, Victor Vaquero, Ahmad El Sallab, Ganesh Sistu, and Senthil Yogamani. Fusedmodnet: Real-time camera and lidar based moving object detection for robust low-light autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019.
- [3] Stephen M. Pizer, E. Philip Amburn, John D. Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B. Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3):355–368, 1987.
- [4] Edwin H. Land. The retinex theory of color vision. *Scientific American*, 237(6):108–129, 1977.
- [5] Heng-Da Cheng and XJ Shi. A simple and effective histogram equalization approach to image enhancement. *Digital signal processing*, 14(2):158–170, 2004.
- [6] Mohammad Abdullah-Al-Wadud, Md Hasanul Kabir, M Ali Akber Dewan, and Oksam Chae. A dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(2):593–600, 2007.
- [7] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013.
- [8] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.
- [9] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021.
- [10] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017.
- [11] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3063–3072, 2020.
- [12] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018.
- [13] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018.
- [14] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1780–1789, 2020.
- [15] Yijun Liu, Zhengning Wang, Yi Zeng, Hao Zeng, and Deming Zhao. Pd-gan: Perceptual-details gan for extremely noisy low light image enhancement. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1840–1844. IEEE, 2021.
- [16] Sos S Agaian, Karen Panetta, and Artyom M Grigoryan. Transform-based image enhancement algorithms with performance measure. *IEEE Transactions on image processing*, 10(3):367–382, 2001.
- [17] Akhilesh Bijalwan, Aditya Goyal, and Nidhi Sethi. Wavelet transform based image denoise using threshold approaches. *International Journal of Engineering and Advanced Technology (IJEAT) ISSN*, 1:2249, 2012.
- [18] Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. Multi-level wavelet-cnn for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 773–782, 2018.
- [19] Ali M Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, 38(1):35–44, 2004.

- [20] Haidi Ibrahim and Nicholas Sia Pik Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(4):1752–1758, 2007.
- [21] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22(12):5372–5384, 2013.
- [22] Keita Nakai, Yoshikatsu Hoshi, and Akira Taguchi. Color image contrast enhancement method based on differential intensity/saturation gray-levels histograms. In *2013 International Symposium on Intelligent Signal Processing and Communication Systems*, pages 445–449. IEEE, 2013.
- [23] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462, 1997.
- [24] Zia-ur Rahman, Daniel J Jobson, and Glenn A Woodell. Multi-scale retinex for color image enhancement. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 3, pages 1003–1006. IEEE, 1996.
- [25] Chang-Hsing Lee, Jau-Ling Shih, Cheng-Chang Lien, and Chin-Chuan Han. Adaptive multiscale retinex for image contrast enhancement. In *2013 International Conference on Signal-Image Technology & Internet-Based Systems*, pages 43–50. IEEE, 2013.
- [26] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997.
- [27] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018.
- [28] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2782–2790, 2016.
- [29] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017.
- [30] Liang Shen, Zihan Yue, Fan Feng, Quan Chen, Shihao Liu, and Jie Ma. Msr-net: Low-light image enhancement using deep convolutional network. *arXiv preprint arXiv:1711.02488*, 2017.
- [31] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021.
- [32] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6849–6857, 2019.
- [33] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018.
- [34] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, page 220, 2018.
- [35] Quan Pan, Lei Zhang, Guanzhong Dai, and Hongai Zhang. Two denoising methods by wavelet transform. *IEEE transactions on signal processing*, 47(12):3401–3406, 1999.
- [36] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [37] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2281–2290, 2020.
- [38] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M Taha, and Vijayan K Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*, 2018.
- [39] Yao Lu, Wei Zhang, Cheng Jin, and Xiangyang Xue. Learning attention map from images. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1067–1074. IEEE, 2012.
- [40] William N Venables and Brian D Ripley. *Modern applied statistics with S-PLUS*. Springer Science & Business Media, 2013.

- [41] Hiroshi Fukui, Tsubasa Hirakawa, Takayoshi Yamashita, and Hironobu Fujiyoshi. Attention branch network: Learning of attention mechanism for visual explanation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10705–10714, 2019.
- [42] Umme Sara, Morium Akter, and Mohammad Shorif Uddin. Image quality assessment through fsim, ssim, mse and psnr—a comparative study. *Journal of Computer and Communications*, 7(3):8–18, 2019.
- [43] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.
- [44] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [45] K. Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2015.
- [46] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.
- [47] Zhenqiang Ying, Ge Li, and Wen Gao. A bio-inspired multi-exposure fusion framework for low-light image enhancement. *arXiv preprint arXiv:1711.00591*, 2017.
- [48] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new low-light image enhancement algorithm using camera response model. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 3015–3022, 2017.
- [49] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new image contrast enhancement algorithm using exposure fusion framework. In *International Conference on Computer Analysis of Images and Patterns*, pages 36–46. Springer, 2017.
- [50] Xutong Ren, Mading Li, Wen-Huang Cheng, and Jiaying Liu. Joint enhancement and denoising method via sequential decomposition. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2018.
- [51] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 129:82–96, 2016.
- [52] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *arXiv preprint arXiv:1706.08500*, 2017.