Enhancement of Low-Light Image Based on Wavelet U-Net

Yuanchen Wang, Xiaonan Zhu, Yucong Zhao, Ping Wang and Jiquan Ma*

Computer Science and Technology, Heilongjiang University, Harbin, Heilongjiang, 150080, China

*Corresponding author’s e-mail: majiquan@hlju.edu.cn

Abstract. In computer vision, low-light image enhancement has always been a challenging task caused by more lower signal to noise ratio. Some methods have been proposed to enhance the low-light image using fully convolution network. Using u-net as backbone, we introduce wavelet transform to conduct down-sampling and up-sampling operations. In order to recover more details, perceptual loss has been used to optimize the network parameters. Experiments show that our model can get better performance than the existing methods. We find that wavelet transform effectively improve the quality of low-light image enhancement.

1. Introduction

Low-light image enhancement has always been a hot point in computer vision, that is useful for many downstream computer vision applications, such as target detection, object tracking and image classification. Images acquired under low-light conditions are affected by illumination or devices. Generally, much more noise will be contained in the signal and underexposed areas of the image are often difficult to detect. In our work, we perform low-light image enhancement using images with severely limited illumination and short exposure.

For low-light image enhancement, a very classic method is histogram equalization [1]. Histogram equalization will lead to the change of brightening area of the whole image, and the noise will increase accordingly. This method usually does not achieve the desired effect on complex scenes. Single scale retinex (SSR) [2] considers images can be decomposed into a pixel-wise product of reflectance and illumination. Multi-Scale Retinex (MSR) [3] also use the illumination map to solve the problem. After a short while, Multi-Scale Retinex with Color Restoration (MSRCR) [4] have been proposed. Fu [5] et al. introduced a weighted vibration model that simultaneously estimates reflectance and illumination. Dong [6] and others were inspired by the dark channel to defog the image, using the high similarity between the low-light image and the dense fog to deal with problem. Guo [7] et al. provided the LIME, taking into account the illumination map estimation and denoising. In recent years, many researchers have proposed to use deep learning to enhance low-light images. Lore [8] et al. came up with using a deep auto-encoder to enhance images in low-light environments. Shen [9] et al. proposed that the MSR-net directly learns the mapping relationship between low-light images and bright images, and then enhanced low-light images. Chen [10] et al. collected a new dataset with an original short exposure and corresponding long-exposure reference images and designed a fully convolutional neural network to solve the problem of image enhancement under extremely low-light condition. Wang, R [11] et al. developed a method of image enhancement under low-light based on deep learning to optimize illumination. The illumination of the intermediate result is learned through a neural network, and then the enhanced image is obtained through the relationship between the input image and the illumination. At present, few people use the wavelet transform to solve low-light problem.
In our work, we propose a new neural network to make full use of the advantages of wavelet transform and perceptual loss [12]. We combine wavelet transform with the u-net to convey and share different frequency signal in low-light image. In addition, we use perceptual loss and $L_1$ loss to train the network parameters. In order to evaluate our method, we conduct a lot of experiments. Experiment results show that our method is superior to other methods.

2. Method
In this section, we will introduce the wavelet transform briefly, our network architecture, loss functions and implementation details. Concrete details will be described in 2.1, 2.2, 2.3 and 2.4.

2.1. Wavelet Transform
An image can be transformed into four sub-images by wavelet transform. The four sub-images are sub-band LL, horizontal detail LH, vertical detail HL, and diagonal detail HH. The resolution of each sub-image becomes half of the original image. In addition, we can find that the noise mainly exists in the high-frequency information. For low-light image enhancement, using u-net alone will not have obvious effects, some details are difficult to learn, so we introduce wavelet transform. We enhance the low-light image by wavelet decomposition and reconstruction, because wavelet transform can get more specific high-frequency information and low-frequency information so that the network can learn more features. We can start with these information to solve the problem.

2.2. Network Architecture
We use the wavelet transform to replace pooling operations in traditional u-net. Our network includes four wavelet transform and four inverse wavelet transform operations. The image after wavelet transform contains rich information, which is beneficial to the network further to learning the details and necessary features. Our network architecture is shown in Figure 1.

![Figure 1. Our network architecture. It includes convolution, wavelet transform, inverse wavelet transform and concat operations.](image)

The pre-processing of Bayer arrays includes: packing the raw image into 4 channels, setting the amplification ratio and subtracting the value of black level, after that, the processed image is sent into our network for learning. After convolution, wavelet transform, inverse wavelet transform and concatenate operations, the output is a 12-channel image, which we converted to RGB format subsequently. There are five layers in our network, using kernel size $3 \times 3$ and LReLU nonlinearity during the convolution operations, with kernel numbers of 32, 64, 128, 256, and 512 respectively.
2.3. Loss Function

We use perceptual loss and $L_1$ loss respectively. In order to better calculate the loss between the prediction image and ground truth, we also combine perceptual loss with $L_1$ loss for training. We cite two expressions of the loss functions, one is the $L_1$ loss and the else is perceptual loss.

$L_1$ loss function is expressed as:

$$L(\hat{y}_i, y_i) = \frac{1}{N} \sum_{i=1}^{N} ||\hat{y}_i - y_i||_1$$  \hspace{1cm} (1)

Where $N$ is the number of all training samples, $\hat{y}_i$ is prediction, and $y_i$ is ground truth. Perceptual loss is expressed as:

$$L(\hat{y}, y) = \frac{1}{C_H C_W} \left\| \Phi_j(\hat{y}) - \Phi_j(y) \right\|_2^2$$  \hspace{1cm} (2)

Where $j$ denotes the $j$th layer of the network, $C_H C_W$ represents the size of the feature map of the $j$th layer, and $\Phi$ is used to denotes the loss network. $\hat{y}$ is prediction, and $y$ is ground truth.

2.4. Implementation Details

We build our network on TensorFlow, and we use perceptual loss and $L_1$ loss. For perceptual loss, we select the feature map from relu2_2 to calculate the loss. In addition, we use Adam optimizer. Since the image resolution of the dataset is relatively large, we randomly crop each image into a 512×512 patch and apply flipping for data augmentation, which is send to the network for training subsequently. Our network train a total of 3500 epochs. The learning rate of the first 2000 epochs is set to $10^{-4}$, and the learning rate of the last 1500 epochs decrease to $10^{-5}$. We try to set the learning rate smaller, but the effect is not significantly improved.

3. Experiment

In this section, we will introduce the training dataset briefly and show the results of the visual comparison and evaluation metrics. It seems that the low value of PSNR or SSIM does not necessarily mean that the results are not good. In other words, we can't just compare the values of the PSNR or SSIM and ignore the visual perception. We conduct a large number of experiments, and results show that our method can improve brightness and remove noise at the same time. Through qualitative and quantitative analysis, it can be concluded that our method is superior to other methods.

3.1. Training Dataset

Many researchers who study low-light image enhancement often use synthesized data. In order to ensure the effectiveness of our method, we used See-in-the-Dark (SID) [10] dataset. The dataset is captured in real scenes by Sony cameras with a resolution of 4240×2832, and it contains indoor images and outdoor images. The short-exposure images were used as input, and the long-exposure images were used as ground truth. The dataset is captured in extremely low light conditions, it is difficult for the naked-eye to distinguish the information of the image. Raw image works better in extremely low-light conditions, and it contains more information.

3.2. Results

(a) (b)
Figure 2. Results enhanced by different methods, and we only use perceptual loss here. (a) input. (b) our result. (c) traditional pipeline. (d) followed by BM3D denoising. (e) u-net.

Figure 3. Results enhanced by our method and u-net method, and we only use perceptual loss here. (a) input. (b) u-net. (c) our result.

| Method        | Loss Function  | PSNR  | SSIM  |
|---------------|----------------|-------|-------|
| u-net         | $L_1$ loss     | 28.88 | 0.787 |
| wavelet u-net | $L_1$ loss     | 29.03 | 0.874 |
| wavelet u-net | perceptual loss| 24.24 | 0.589 |
| wavelet u-net | perceptual loss + $L_1$ loss | 28.92 | 0.851 |

4. Discussion
In our work, we find that some enhanced images were overexposed, and what we want to do is not changing the normal exposure area of the image. In addition, we also find that visual effects with perceptual loss participation are better than without participation, but the value of the PSNR or SSIM is not high, we will continue to explore the reasons for this result. We hope that the current work can bring inspiration to other researchers.
5. Conclusion
We propose a novel method for enhancing low-light images. Using u-net as backbone, we introduce wavelet transform. And perceptual loss and $L_1$ loss have been used to optimize the network parameters for recover more details. We compare our wavelet u-net with traditional u-net from the qualitative and quantitative perspective respectively, the results show that our method has obvious effects. In the future, we will continue to explore better network architecture and new loss functions to further optimize our results. Further, we will apply them to low-light video enhancement and other computer vision tasks.

References
[1] Singh, R. P., & Dixit, M. (2015) Histogram equalization: a strong technique for image enhancement. International Journal of Signal Processing, Image Processing and Pattern Recognition., 8(8): 345-352.
[2] Jobson, D. J., Rahman, Z. U., & Woodell, G. A. (1997) Properties and performance of a center/surround retinex. IEEE transactions on image processing., 6(3): 451-462.
[3] Rahman, Z. U., Jobson, D. J., & Woodell, G. A. (1996) Multi-scale retinex for color image enhancement. In: Proceedings of 3rd IEEE International Conference on Image Processing. Vol. 3, pp. 1003-1006.
[4] Jobson, D. J., Rahman, Z. U., & Woodell, G. A. (1997) 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.
[5] Fu, X., Zeng, D., Huang, Y., Zhang, X. P., & Ding, X. (2016) A weighted variational model for simultaneous reflectance and illumination estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2782-2790.
[6] Dong, X., Wang, G., Pang, Y., Li, W., Wen, J., Meng, W., & Lu, Y. (2011) Fast efficient algorithm for enhancement of low lighting video. In: 2011 IEEE International Conference on Multimedia and Expo. pp. 1-6.
[7] Guo, X., Li, Y., & Ling, H. (2017) LIME: Low-Light Image Enhancement via Illumination Map Estimation. IEEE Trans. Image Processing., 26(2): 982-993.
[8] Lore, K. G., Akintayo, A., & Sarkar, S. (2017) LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition., 61: 650-662.
[9] Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., & Ma, J. (2017) Msr-net: Low-light image enhancement using deep convolutional network. arXiv preprint arXiv:1711.02488.
[10] Chen, C., Chen, Q., Xu, J., & Koltun, V. (2018) Learning to see in the dark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3291-3300.
[11] Wang, R., Zhang, Q., Fu, C. W., Shen, X., Zheng, W. S., & Jia, J. (2019) Underexposed Photo Enhancement Using Deep Illumination Estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6849-6857.
[12] Johnson, J., Alahi, A., & Fei-Fei, L. (2016) Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. Springer, Cham. pp. 694-711.