Feature Fusion Network for Low-Light Image Enhancement

Yabin Yu*

1 College of Computer Science, Sichuan University, Chengdu, Sichuan, 610065, China
*yuyabin@stu.scu.edu.cn

Abstract. With the popularity of electronic cameras, we can get pictures under various lighting conditions in our daily life. The images which obtained in low illumination environment will lose contrast, colour, and content details which will bring obstacles for understanding the image. To tackle this problem, a method is proposed in this paper to tackle low-light image enhancement. This method fuses the high level and the low level feature, uses an attention module to model the context information and selects the critical feature to supplemented the high level feature to reconstruct the enhanced image. A number of experiments demonstrate that our method can get images with improved and reduced noise.

1. Introduction

In the real world, we often get uniform lighting images due to the limitation of environment or the limited equipment. Images which obtained under the insufficient illumination often has the problem such as the colour distortion and missing details which will cause misunderstanding of the image content. Methods to recover the low light images are very essential.

Many researchers have proposed many solutions to enhance low-light images. The HE-based method[1] just stretch the pixel values of low-light images to obtain brighter images. These methods can enhance the contrast of the images in a simple way, but they will lose image details. Retinex theory[2] can be used to get the enhanced images. Many methods have proposed based on the theory. From the SSR[3] to MSRCR[4] and other. LIME[5] estimates the illumination map based on the Retinex theory. Although the above methods can recover the brightness of the image, they get unnatural results.

In recent years, convolutional neural network has made great progress in various computer vision tasks. Methods based on CNN also have been proposed. Some methods combine the CNN with the Retinex theory[6-7].

The U-Net[8] network has been used in various computer vision task and obtain great success. And there are enhancement method[9] also use the form of the U-Net. The method of [9] which process the RAW images is not suitable for many situations. The U-Net can get multi contextual information at the downsampling stage and obtain the feature with semantics information at the high stage. This paper proposes a method which using a context attention module to perceive the context information of the low-level features and learn to choose the critical features which can supplement the missing details in the reconstruction stage under the guidance of high-level features. Experiments has made to validate the effective of the network.

2. Proposed Method

The detail of the network proposed in this paper will be given and description of how the high stage feature guide the attention module to learn the feature for enhancement from the low-stage features in this section.
2.1. Network Architecture
The network proposed in this paper as Figure 2 shows. This paper uses a U-Net to enhance the low-light images for end-to-end. Firstly, this method downsamples the images with four stride convolutions, at this stage the features have spatial details such as the edge information and context information. At the upsampling stage, it uses the deconvolution layer and one convolution layer as one block. After upsampling the feature map through deconvolution layers, the features will fuse the corresponding feature of the downsampling features. And the fusion features will feed to the attention module. And the output of the attention module will add the origin high level feature. The context attention module learns the missing details from the low-stage feature with the guidance of the features of the corresponding upsampling stage.

![Network Architecture Diagram]

Figure 1 The network architecture proposed in this paper.

2.2. Attention module
There are many computer vision tasks which used the attention mechanism has proved the effectiveness of the attention mechanism. This method uses the context attention module from [10]. As we all know, the network does not make a distinction between each feature map and has a limited receptive filed which fails to capture the long-range information. The above two deficiency of the network can be solved by the SE block[11] and non-local block[12] separately.

This paper uses a context attention module to learn to perceive the context information from the low level feature. And with the guide of the information of the high level features, the attention module can learn what the feature of the low level is critical for high level features in the upsampling stage. The attention module used in this paper has the two advantages of the SE block and the non-local block with less calculation burden.
The contextual attention model which is illustrated in Figure 2 can be defined as

\[ y_j = x_j + \delta \left( \sum_{j=0}^{N} \frac{e^{w_j}}{\sum_{m=0}^{N} e^{w_m}} x_j \right) \]  

Firstly, the \( \sum_{m=0}^{N} e^{w_m} \) which refers to the context modeling. The \( \delta() \) presents the transform which can capture channel-wise dependencies. Finally, the context feature and the features of each position are fused together.

2.3. Loss function
The pixel-level loss \( L_1 \), the perceptual loss [13] and relativistic average LS adversarial loss [14] are used in this paper.

\[ L_{\text{total}} = \lambda_1 * L_1 + \lambda_2 * L_{\text{adv}} + \lambda_3 * L_{\text{perc}} \]  

The \( \lambda_1, \lambda_2, \lambda_3 \) in this paper are parameters and empirically set as \( \lambda_1 = 1, \lambda_2 = 0.5, \lambda_3 = 0.2. \)

The \( L_1 \) loss makes the enhanced images consistent with the ground truth in brightness and colour.

\[ L_1 = ||I_{\text{out}} - I_{\text{gt}}||_1 \]  

To make the output results similar to the ground truth in feature level and make the results look more natural. The perceptual loss \( L_{\text{perc}} \) is employed.

\[ L_{\text{perc}} = \sum_{i} \frac{1}{N_i} \| Q_i(I_{\text{out}}) - Q_i(I_{\text{gt}}) \|_1 \]  

The \( N_i \) is the size of the \( Q_i \) which is the output of the \( i \)-th layer of the pretrained VGG-16.

The network also seen as a generator, and the global and local discriminator from [15] are also used. And the generator adopts the \( L_{\text{adv}} \) as the loss:

\[ L_{\text{adv}} = E_{x,y} [(D_{\text{adv}}(x,y) - 1)^2] + E_{x,y} [D_{\text{adv}}(x,y)^2] \]  

3. Experiments
In this section, subjective evaluation and objective evaluation will be conducted. The methods: SRIE[16], GLADNet[17], LIME[5], and NPE[18] are selected to compare this method. The evaluate index of PSNR and SSIM[19] is chosen for objective evaluation.

This paper uses the dataset from[6] to train the network. The dataset has 500 pairs of low/normal-light image pairs of real-world scenes and 1000 composite images.
3.1. Subjective Evaluation
In this section, the subjective evaluation experiment are conducted on the VV[20]、LIME[5]、MEF[21] datasets.

![Subjective Evaluation Images]

Figure 3 The subjective visual comparison

3.2. Objective Evaluation
This paper takes the objective evaluation experiments on 50 synthetic images. The results are illustrated in the table, and the conclusion that the method proposed in this paper gets a better result than other methods can be drawn from the table below.

| Method   | PSNR  | SSIM  |
|----------|-------|-------|
| SRIE     | 11.86 | 0.50  |
| LIME     | 17.18 | 0.60  |
| NPE      | 17.29 | 0.67  |
| GLADNet  | 19.71 | 0.70  |
| Ours     | 22.03 | 0.72  |

Table 1 The objective evaluation.

4. Conclusions
This paper proposes an effective way which fuses the low-level features and high level-features and a context attention module is adopted to learn the detail features what the upsampling stage needs to enhance the low-light images. The network enhances the low-light images with brightness and reduced noise. However, there are some still shortcomings to improve. The efficiency can be improved by making the network lightweight.

References
[1] Pizer, S.M., Johnston, R.E.(1990) Contrast limited adaptive histogram equalization: speed and effectiveness. In: Proceedings of the First Conference on Visualization in Biomedical Computing.
[2] Land, E.H.(1977) The Retinex Theory of Color Vision. Scientific American, 237(6):108-128.
[3] Jobson, D.J., Rahman, Z. Woodell, G.A.(1997) Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing, vol. 6, no. 3, pp. 451-462, March.
[4] Jobson, D.J., Rahman, Z. and Woodell, G. A. (1997) A multiscale retinex for bridging the gap between color images and the human observation of scenes, in IEEE Transactions on Image Processing, vol. 6, no. 7, pp. 965-976, July.
[5] Guo, X., Yu, L., Ling, H. (2016) Lime: low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, PP(99), 1-1.
[6] Wei, C., Wang, W., Yang, W., Liu, J.(2018) Deep retinex decomposition for low-light enhancement. BMVC 2018, Oral Presentation.
[7] Wang, R., Zhang, Q., Fu, C. W., Shen, X., J, Jia. (2019) Underexposed Photo Enhancement Using Deep Illumination Estimation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.
[8] Ronneberger, O., Fischer, P., Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing.
[9] Zhang, C., Yan, Q., Y Zhu, Li, X., Sun, J., Y Zhang. (2020). Attention-based network for low-light image enhancement. 2020 IEEE International Conference on Multimedia and Expo (ICME). IEEE.
[10] Y, Cao., Xu, J., Lin, S., Wei, F., & Hu, H. (2020) GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). IEEE.
[11] Jie, H., Li, S., Gang, S.(2018) Squeeze-and-Excitation Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.
[12] Wang, X., Girshick, R., Gupta, A., He, K. (2017) Non-local neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7794–7803.
[13] Johnson, J., Alahi, A., Fei-Fei, L. (2016) Perceptual losses for real-time style transfer and super-resolution. Springer, Cham.
[14] Mao, X., Li, Q., Xie, H., Lau, R., Smolley, S.P. (2017) Least squares generative adversarial networks. 2017 IEEE International Conference on Computer Vision (ICCV).
[15] Jolicoeur-Martineau, A..(2018). The relativistic discriminator: a key element missing from standard gan..arXiv preprint arXiv:1807.00734.
[16] Fu, X., Zeng, D., Yue, H., Zhang, X. P., Ding, X.(2016) A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE.
[17] Wang, W., Chen, W., Yang, W., Liu, J.(2018) GLADNet: Low-Light Enhancement Network with Global Awareness. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE Computer Society.
[18] Wang, S., Zheng, J., Hai-Miao, Li, H.(2013) Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. IEEE Transactions on Image Processing, vol. 22, no. 9, pp.3538-3548.
[19] Zhou, W., Bovik, A. C., Sheikh, H.R., Simoncelli, E. P. (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process, 13(4).
[20] Vonikakis,V.https://sites.google.com/site/vonikakis/dataset/challenging-dataset-for-enhancement.
[21] Ma, K., Zeng, K., Wang, Z. (2015) Perceptual quality assessment for multi-exposure image fusion. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 24(11), 3345.