Image Enhancement of Low Light UAV via Global Illumination Self-aware feature Estimation

Jingyu Niu
School of Telecommunication and Information Engineering
Xi'an University of Posts and Telecommunications
Xi'an 710121, China
1398473442@qq.com

Dianwei Wang*
School of Telecommunication and Information Engineering
Xi'an University of Posts and Telecommunications
Xi'an 710121, China
wangdianwei@xupt.edu.cn

Pengfei Han
School of Telecommunication and Information Engineering
Xi'an University of Posts and Telecommunications
Xi'an 710121, China
13772405149@126.com

Jie Fang
School of Telecommunication and Information Engineering
Xi'an University of Posts and Telecommunications
Xi'an 710121, China
jackfang713508@gmail.com

Xincheng Ren
School of Physics and Electronic Information
Yan'an University
Yan'an, 716000,China
xchren@yau.edu.cn

Yongrui Qin
School of Computing and Engineering
University of Huddersfield
Huddersfield, HD1 3DH, UK
Yongrui.Qin@hud.ac.uk

Zhijie Xu
School of Computing and Engineering
University of Huddersfield
Huddersfield, HD1 3DH, UK
z.xu@hud.ac.uk

***Abstract*—UAV images acquired under low light conditions are often characterized by low contrast and poor visual effect. To improve image quality, a low light UAV image enhancement method via global illumination self-aware feature estimation was proposed. First, a novel lightweight GhostNet is introduced to extract deeper image features. Secondly, the self-aware module is used to correct the possible missing information between encoder network and decoder network. Finally, gradient loss and structural similarity loss are used to constrain the network to achieve the goal of edge preservation and detail restoration. Through extensive experiments, the method proposed can effectively improve the visualization effect, and get more natural and real results.**

Keywords—image enhancement, UAV image, feature extraction, GhostNet, self-aware

I. INTRODUCTION

Unmanned aerial vehicle (UAV) is widely used in society, such as reconnaissance, security maintenance, disaster relief. However, due to environmental factors, the obtained UAV image is difficult to be applied directly. So the low light UAV image enhancement has become a hotspot in computer vision [1,2].

Over the past decades, with the continuous development of low light image enhancement, many researchers turn their attention from the traditional model-driven method to the data-driven method, and try to link them together, so the data-driven is no longer a pure black box. For example, the decomposition idea of the Retinex-class enhancement algorithm gives great inspiration to data-driven, Shen et al. [3] believed that the traditional MSR method could be regarded as a convolutional neural network with different Gaussian convolution kernel, and proposed MSRNet to learn the end-to-end mapping of images. Wang et al. [4] constructed a global illumination evaluation network, which first estimated illumination and refined it. Park et al. [5] proposed a dual self-coding network based on Retinex. Wei et al. [6] proposed RetinexNet after processing the reflection and illumination components separately. Inspired by it, Zhang et al. [7] considered the impact of noise degradation and used illumination to guide reflection recovery and proposed the network named KIND. Wang et al. [8] designed a kind of progressive Retinex framework to make illumination and noise enhance each other mutually. Xu et al. [9] proposed an enhanced model based on frequency decomposition. In general, because the UAV image has the characteristics of both distant view and close shot, the existing methods can not apply for the UAV image enhanced task.

In order to solve this problem, we proposed a low light UAV image enhanced method. Specifically, to understand the image information efficiently, the encoder structure is composed of ghost module [10] nested by Squeeze-and excitation (SE) module [11], ghost module as a lightweight feature extraction module, which fuses the deep features containing semantic content and the shallow features containing texture details. A non-local operation [12] is inserted in the encoder-decoder network to obtain the dependence on the long-distance and wide-range of the image, it further compensates for the details that may be lost in the scaling process. In addition, we use the loss function containing gradient information for edge preservation, and suppress the artifacts. A large number of experiments show that the proposed method can obtain better visuals.

The main contributions of this paper are as follows:

1) We proposed a low light UAV image enhanced method via global illumination self-aware feature estimation, which can effectively improve the quality of UAV images.

2) To capture image information better while reducing unnecessary computation, a simplified ghost feature extraction network is designed, and a self-aware module is introduced to integrate local and non-local information.

2021 3rd International Conference on Natural Language Processing (ICNLP) | 978-1-6654-1411-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICNLP52887.2021.00044

978-1-6654-1411-1/21/$31.00 ©2021 IEEE
DOI 10.1109/ICNLP52887.2021.00044

3) Compared with several existing state-of-art methods, the proposed method can get ideal results with various details.

The remaining sections of this paper are organized as follows. Section 2 introduces our method in detail. Section 3 gives the comparison results from three aspects. Section 4 conducts this paper.

II. THE PROPOSED METHOD

In order to recover more detailed information, a self-aware global illumination feature estimation method for low light UAV image enhancement was proposed. It consists of two steps: the first step is global illumination estimation based on self-aware, we use a lightweight feature extraction module nested by non-local operations to construct the encoder-decoder network, and the illumination component is estimated at the bottleneck layer. The second step is detail restoration, the convolutional network combined with the input image is used to supplement the information that may be lost in the previous process, it can be enhanced more accurately under the reference of the estimated illumination. The proposed framework is shown in Fig. 1.

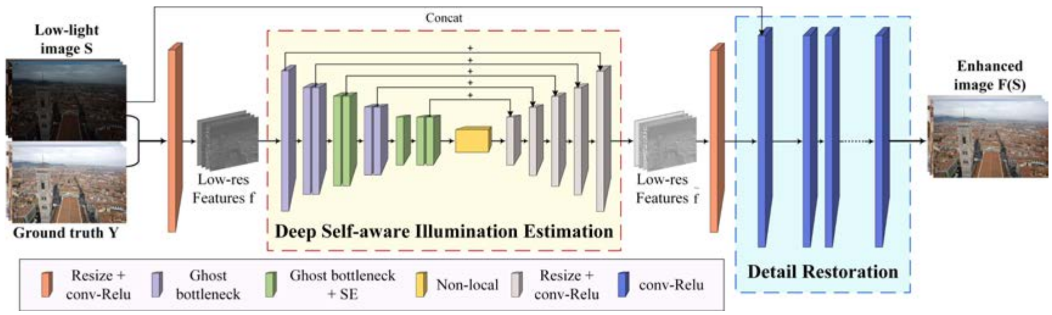

Fig. 1. The proposed framework.

A. Self-aware Global Illumination Estimation

The self-aware global illumination estimation is released by an encoder-decoder network, it consists of an encoder network capturing each level's feature and details, a non-local operation module getting the context information, and a decoder network reconstructing the content features and vector dimensions. To be specific, the framework is divided into three steps: first, the input image is scaled to a feature map with fixed resolution. Second, feed the feature map to the encoder-decoder network. Last, the global illumination is estimated at the bottleneck layer with the receiving field covering the whole image.

Consider to minimize the unnecessary computation while getting deeper features, we build an improved lightweight feature extraction network, which has fewer layers than GhostNet. According to the requirements, we design a simpler input resolution $W_0 \times H_0$ for the network, then construct the number and structure of ghost bottleneck blocks. To make the network be able to calibrate the features effectively, some SE modules are inserted into the ghost bottleneck blocks to extract the local channel features.

$$F_{enc} = g_6'\left(g_5'\left(g_4\left(g_3'\left(g_2\left(g_1\left(R(S)\right)\right)\right)_\downarrow\right)_\downarrow\right)_\downarrow\right)_\downarrow\right)_\downarrow \tag{1}$$

where F_{enc} is the output of encoder block, R is the resize operation, g_i is ghost bottleneck, g_i' is ghost bottleneck with channel attention module, stride size of g_1 and g_6' is set to 1, while the stride of other modules is set to 2. $\downarrow$ denotes the down sampling operation.

The input feature map is feed to the non-local operation unit through a series of ghost bottleneck blocks. Since the paired UAV images share the same texture and colors, they could provide a lot of context prior information for correcting the uneven illumination area. Moreover, adding non-local units can expand the receiving field to the whole image and capture the remote dependence. Therefore, the combination of non-local and local information can build a richer feature representation.

$$F_{nlu} = NLU(F_{enc}) \tag{2}$$

where NLU represents a non-local operation.

Finally, We input the self-aware features into the convolutional layers for decoding, and the hierarchical feature representation is further utilized to learn the residual information. In a word, the improved network can improve network performance.

$$F_{dec} = \left(f_5\left(f_4\left(f_3\left(f_2\left(f_1(F_{nlu}), F_{\text{enc}}\right)_\uparrow, g_4\right)_\uparrow, g_3'\right)_\uparrow, g_2\right)_\uparrow, g_1\right)_\uparrow \tag{3}$$

where F_{dec} is the result of encoder-decoder network output, f_i is the convolution function, $\uparrow$ denotes the up-sampling operation, and $(,)$ denotes the skip connection by channel.

B. Detail Recovery Network

Because the information may be lost during scaling or up-down sampling in the previous step, and the input image contains the most detailed structure, so we use the detail recovery network to make up for this shortcoming. Specifically, we connect the input and the feature map, so the enhanced image retains the illumination estimation and richer information entropy, which is more in line with our requirements.

$$F_{res} = f_6\left(R\left(F_{dec} \oplus S\right)\right) \tag{4}$$

where F_{res} is the enhanced result after detail recovery, and $\oplus$ is CONCAT operation. This network is composed of a convolutional layer with 5 ReLU activation functions.

C. Loss Function

The loss function uses one or several constraints to represent the difference between the predicted and the actual data, and help us to get the optimal mapping parameters from the enhanced image $F(X,\Theta)$ to the reference image Y. Therefore, the loss function in this paper is:

$$L = \|Y - F(S)\|_2^2 + \|\nabla Y - \nabla F(S)\|_2^2 + \mathrm{SSIM}(Y, F(S)) \quad (5)$$

where $\mathrm{SSIM}(.)$ is the structural similarity measure, Y is ground truth, $F(S)$ is proposed operation.

III. EXPERIMENTAL COMPARATIVE

A. Dataset

We select the 400*600 synthetic data set made by GladNet as the training data. 780 original images from RAISE [13] are used as training data, and 80 pieces to verify. In addition, in a real low light environment, 15 pictures were taken by DJI Sprit series UAV as the test data set of this experiment. In order to verify the robustness of the experimental results, we also test on SICE [14] and other data sets.

B. Experiments Settings

We trained our network 50 epochs, which used TensorFlow on a PC with a NVIDIA Titan XP×4 GPU and Intel Xeon e5-2620v4 × 2 CPU. The proposed method uses 3*3 convolution kernel and bottleneck size set by GladNet to obtain a larger receiving field. We use the initialization method proposed in [15] to initialize the weight and adopted Adam optimization. The initial learning rate is set to 0.001.

C. Subjective Evaluation

In order to verify the effectiveness of this method, we choose several existing state-of-the-art methods, such as dong [16], LIME [17], RetinexNet [6], GLAD [4], KIND [7] for comparison. 6 low light UAV images of different scenes were selected for experiments, and the experimental results as shown in Fig. 3~7.

Fig. 2. Enhanced results of different method

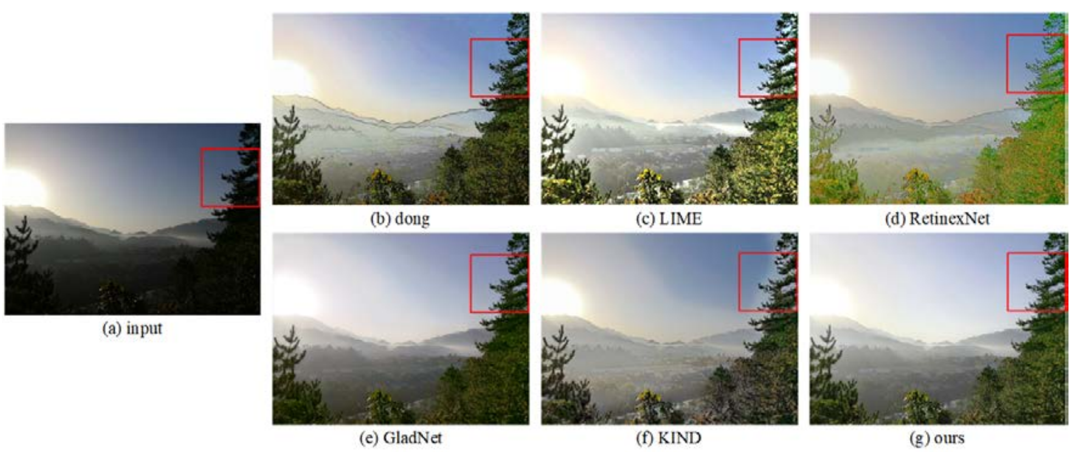

Fig. 3. Enhanced results of different method

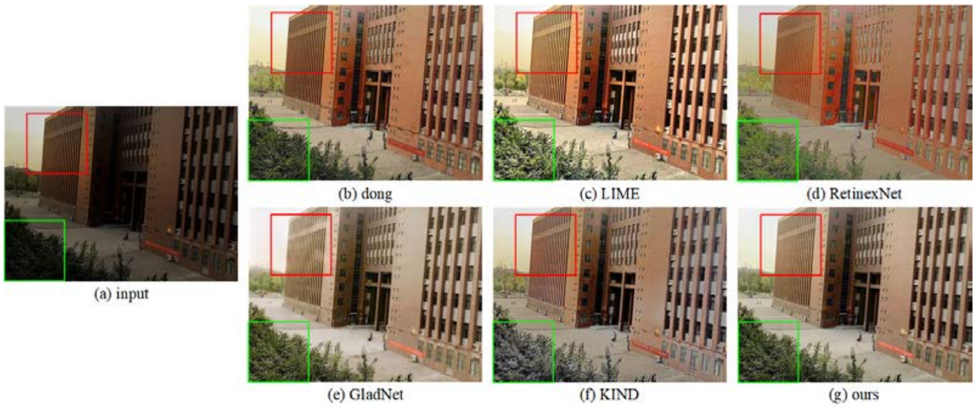

Fig. 4. Enhanced results of different method

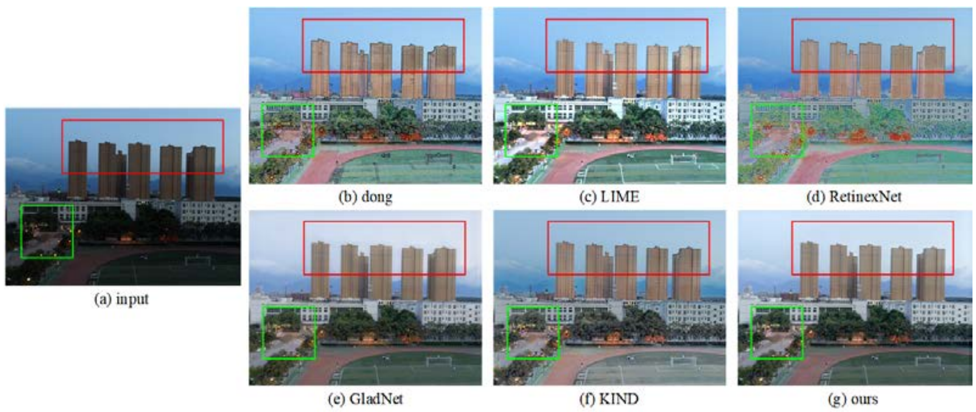

Fig. 5. Enhanced results of different method

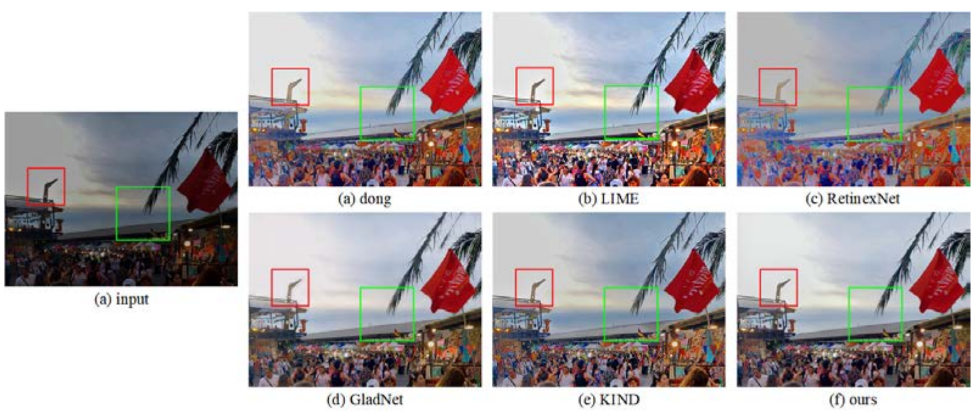

Fig. 6. Enhanced results of different methods

Fig. 2~Fig. 6 show the enhanced results in several different scenes. Dong's method has certain limitations that results show obvious content missing. When an image contains uneven illumination areas, LIME makes enhanced results over-enhanced and detail loss in Fig. 3~5. RetinexNet's results have a poor visual effect, it can't recover the correct content or color, as shown in Fig. 2. GladNet improves brightness better, but the results sometimes have edge blur, color leakage and halo phenomenon, as shown in Fig. 4 and Fig. 6. KIND strengthens the edge information to a certain extent, but the color distortion has not been significantly solved, as shown in Fig. 3. Compared with other methods, our proposed method can effectively alleviate color leakage and artifacts. In other words, it can retain more details in the dark and make the results more in line with what the human eye sees as the real scene.

To further prove that our method has a better visual effect, we test it on the SICE dataset. As shown in Fig. 7 and Fig. 8, the comparison between the enhanced methods and ground truth are listed.

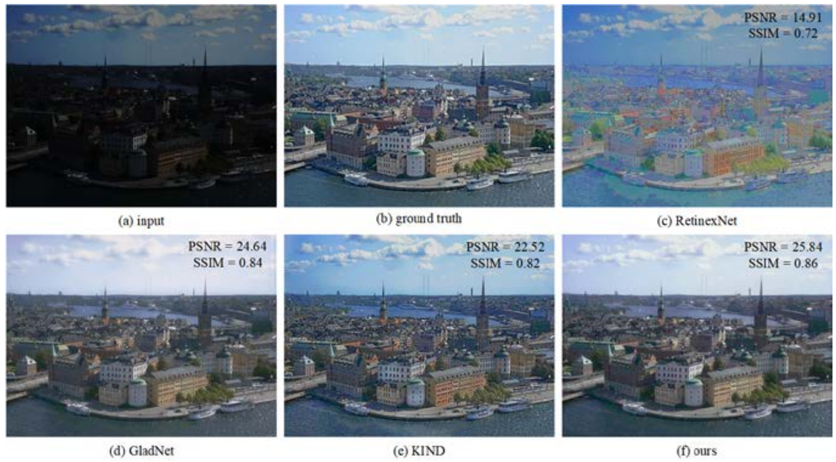

Fig. 7. Enhanced results of different methods.

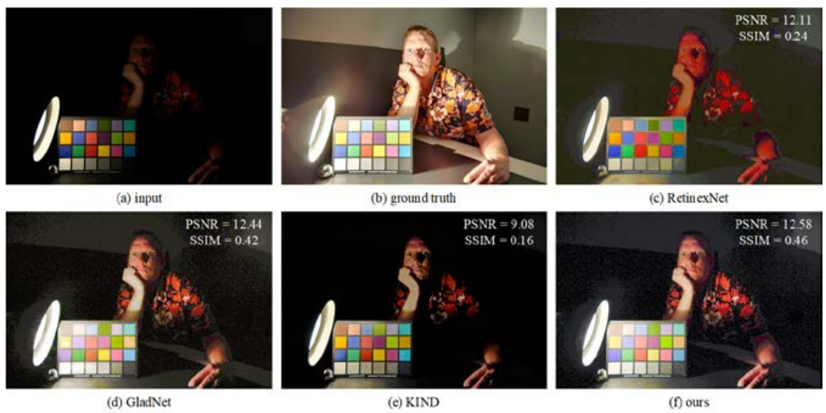

Fig. 8. Enhanced results of different methods.

TABLE I. COMPARISON OF OBJECTIVE EVALUATION RESULTS

		Figure 2	Figure 3	Figure 4	Figure 5	Figure 6
dong	LOE	577.70	451.10	1329.00	1118.00	703.90
	AG	6.53	6.50	14.47	7.24	9.78
	NIQE	4.26	2.88	2.93	**2.91**	3.44
	BRISQUE	36.69	17.08	39.60	28.49	41.19
LIME	LOE	1544.00	956.40	1266.00	1304.00	1505.00
	AG	9.02	5.06	14.55	8.40	11.83
	NIQE	4.52	2.90	2.92	2.92	2.81
	BRISQUE	37.52	15.39	42.54	28.68	40.13
RetinexNet	LOE	1229.00	969.90	2330.00	1497.00	2163.00
	AG	7.50	8.26	13.56	10.06	9.41
	NIQE	**4.24**	3.25	5.25	5.25	3.26
	BRISQUE	**17.06**	9.55	39.51	31.26	**36.89**
GladNet	LOE	206.60	1820.00	436.50	**316.50**	471.50
	AG	9.78	8.50	16.22	11.88	12.24
	NIQE	5.12	3.18	2.60	3.84	2.88
	BRISQUE	40.08	11.65	44.05	28.31	41.23
KIND	LOE	294.50	381.70	964.00	789.50	501.00
	AG	4.36	2.62	7.16	4.80	**15.64**
	NIQE	5.48	3.70	3.03	3.56	2.95
	BRISQUE	34.20	17.51	36.86	36.17	46.36
ours	LOE	**162.00**	**320.60**	**278.90**	327.00	**425.40**
	AG	**13.88**	**10.09**	**21.83**	**12.53**	9.96
	NIQE	4.99	**2.82**	**2.34**	3.79	**2.80**
	BRISQUE	29.87	**8.98**	**38.95**	**20.77**	40.57

As can be seen in Fig. 7, Compared with input image, RetinexNet and KIND make experiment results lose detail. When the GladNet is applied to the distant view, it usually produces artifacts and the outline of the object is blurred. All three of these methods cause color distortion in Fig. 8, In general, the proposed methods can better distinguish between dark and light colors, and the restored colors are closest to the ground truth.

D. Objective Evaluation

We use Lightness Order Error (LOE) [18], Average Gradient (AG), Natural Image Quality Evaluator (NIQE) [19] and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [20] to evaluate the results, as shown in Table.1. The larger the AG value is, the clearer the image is, while the smaller values the other metrics are, the better the image quality is.

Table 1 shows the comparison of the experimental results of four quantitative metrics for different methods. Based on the above four quantitative metrics, our method has better quantitative metrics. It can keep the image color more natural, and highlight the details in the dark region of the enhanced result.

E. User Study

In order to further verify the superiority of the proposed method, we established a more objective and impartial user. First, 25 low light images were randomly selected from the self-built UAV dataset, DICM [21], LIME [17], MEF [22] and NPE [23] datasets. Then, each image was enhanced in 6 different methods, and 25 groups of 150 enhanced images were obtained.

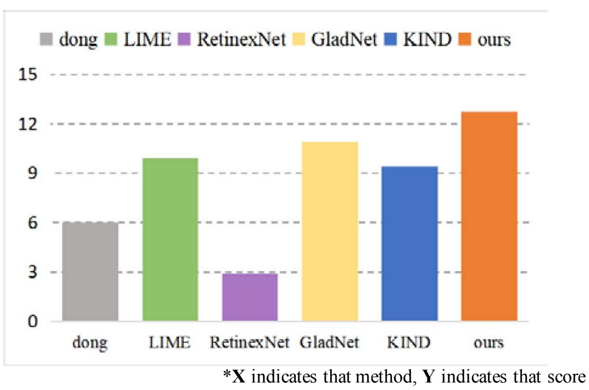

Fig. 9. User study.

We invited 10 subjects to evaluate the results of each group of images according to their preferences, and used 1 to 6 (worst to best) to score via the following criteria: (1) There is no phenomenon of oversaturation or overexposure. It is required that the resulting image should conform to common sense, and be the actual effect that human eyes can see in nature. (2) There are no obvious artifacts, such as halo, shadows, and color leakage of enhanced results. (3) Enhanced results should maintain good edge information and clearly distinguish the boundaries of various objects. After the subjects had scored all the selected images, we summarized the scores in Fig. 9.

IV. Conclusion

In this paper, we propose a low light UAV image enhanced method via global illumination self-aware feature estimation. The framework consists of three parts: deep feature integration network, non-local operation module for extracting context information, and detail restoration module. Extensive experiments show that the proposed method can effectively improve the visual quality, and make enhanced images more realistic and natural.

Although the proposed method has a good balance between near and far objects, it cannot well enhance images with extremely dark scenes or extremely uneven illumination. In the future, we will focus on assigning weights to local areas to see more details.

Acknowledgement

This work was supported in part by the Special Project of Strengthening Police with Science and Technology of Ministry of public security under Grant Nos. 2019GABJC42 and the Fund of Shaanxi Key Laboratory of Intelligent Processing for Big Energy Data, Yan'an University under Grant Nos.IPBED6, and the graduate Innovation Fund of Xi'an University of Posts and Telecommunications No. CXJJLY202026.

References

[1] Wang Dian-Wei, Han Peng-Fei, Fan Jiu-Lun, Liu Ying, Xu Zhi-Jie, Wang Jing. Multispectral image enhancement based on illuminance-reflection imaging model and morphology operation. Acta Physica Sinica, 2018, 67(21): 210701.

[2] Han P, Wang D, Yang X, et al. An Improved Adaptive Correction Algorithm for Non-uniform Illumination Panoramic Image[C]// 2019 IEEE 2nd International Conference on Electronic Information and Communication Technology (ICEICT). IEEE, 2019.

[3] Shen L, Yue Z, Feng F, et al. Msr-net: Low-light image enhancement using deep convolutional network[J]. arXiv preprint arXiv:1711.02488, 2017.

[4] Wang W, Wei C, Yang W, et al. GLADNet: Low-light enhancement network with global awareness[C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG2018). IEEE, 2018: 751-755.

[5] Park S, Yu S, Kim M, et al. Dual autoencoder network for retinex-based low-light image enhancement[J]. IEEE Access, 2018, 6: 22084-22093.

[6] Wei C, Wang W, Yang W, et al. Deep retinex decomposition for low-light enhancement[J]. arXiv preprint arXiv:1808.04560, 2018.

[7] Zhang Y, Zhang J, Guo X. Kindling the darkness: A practical low-light image enhancer[C]//Proceedings of the 27th ACM International Conference on Multimedia. 2019: 1632-1640..

[8] Wang Y, Cao Y, Zha Z J, et al. Progressive Retinex: Mutually Reinforced Illumination-Noise Perception Network for Low-Light Image Enhancement[C]//Proceedings of the 27th ACM International Conference on Multimedia. 2019: 2015-2023.

[9] Xu K, Yang X, Yin B, et al. Learning to Restore Low-Light Images via Decomposition-and-Enhancement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2281-2290.

[10] Han K, Wang Y, Tian Q, et al. GhostNet: More features from cheap operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 1580-1589.

[11] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[12] Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7794-7803.

[13] Dang-Nguyen D T, Pasquini C, Conotter V, et al. Raise: A raw images dataset for digital image forensics[C]//Proceedings of the 6th ACM Multimedia Systems Conference. 2015: 219-224.

[14] Cai J, Gu S, Zhang L. Learning a deep single image contrast enhancer from multi-exposure images[J]. IEEE Transactions on Image Processing, 2018, 27(4): 2049-2062.

[15] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1026-1034.

[16] Dong X, Wang G, Pang Y, et al. Fast efficient algorithm for enhancement of low lighting video[C]//2011 IEEE International Conference on Multimedia and Expo. IEEE, 2011: 1-6.

[17] Guo X, Li Y, Ling H. LIME: Low-light image enhancement via illumination map estimation[J]. IEEE Transactions on image processing, 2016, 26(2): 982-993.

[18] Land E H. The retinex theory of color vision[J]. Scientific american, 1977, 237(6): 108-129.

[19] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.

[20] Mittal A, Soundararajan R, Bovik A C. Making a “completely blind” image quality analyzer[J]. IEEE Signal processing letters, 2012, 20(3): 209-212.

[21] Lee C, Lee C, Kim C S. Contrast enhancement based on layered difference representation[C]//2012 19th IEEE International Conference on Image Processing. IEEE, 2012: 965-968.

[22] Ma K, Zeng K, Wang Z. Perceptual quality assessment for multi-exposure image fusion[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3345-3356.

[23] Wang S, Zheng J, Hu H M, et al. Naturalness preserved enhancement algorithm for non-uniform illumination images[J]. IEEE Transactions on Image Processing, 2013, 22(9): 3538-3548.