Low light enhancement CNN Network based on attention mechanism

Xiwen Liang[1], Xiaoning Yan[2], Nenghua Xu[2], Xiaoyan Chen[1], Hao Feng[1]

[1]*Tianjin University of Science and Technology ,300222, China;*

[2]*Shenzhen Softsz Co.Ltd, Shenzhen, China;*

E-mail: 1540227453@qq.com

www.tust.edu.cn

Abstract

Low-light enhancement is a challenging task. With the image brightness increasing, the noises are amplified, and with the contrast and detail increasing, the false information is generated. In order to solve this problem, this paper proposes a novel end-to-end attention-guided method (A-MBLLEN) based on multi-branch convolutional neural network. The proposed network is composed with enhancement module (EM) and Convolutional Block Attention Module (CBAM). The attention module can make the CNN network structure gradually focus on the weak light area in the image, and the enhancement module can fully highlight the multi-branch feature graph under the guidance of attention. In this manner, image quality is improved from different aspects. Extensive experiments demonstrate that our method can produce high fidelity enhancement results for low-light images quantitatively and visually.

Keywords: Low-light image Enhancement, Deep learning, multi-branch Fusion, Convolutional Neural Network

1. Introduction

Due to unavoidable environments or technical limitations, many photographs are often taken under less than ideal lighting conditions. Poorly litght photos are not only bad for aesthetic quality, but also bad for messaging. The former affects the audience's experience, while the latter leads to misinformation being communicated. To solve these degradation and convert low-quality low light level images into normal light high-quality images, it is necessary to develop a good enhancement technology for low light image. In this paper, a deep neural network structure is proposed to improve the objective and subjective image quality. Extensive experiments demonstrate that our method can produce high fidelity enhancement results for low-light images quantitatively and visually. Our contributions are summarized as follows. 1) We introduce the convolution block attention module (CBAM) into the multi branch enhancement module, which can better highlight the low illumination area. 2) Our method is also effective in suppressing image noise and artifacts in low light region.

2. Related Work

Recently, deep learning has achieved great success in the field of low-level image processing. Powerful tools such as end-to-end networks and GANs[1] have been used in image enhancement. LLNet[2] ues the multilayer perception auto-encoder for low-light image enhancement and denoising. Retinex-Net[3] combines the Retinex theory with CNN to estimate the illumination map and enhance the low-light images by adjusting the illumination map. In order to integrate the advantages of CNN and GAN, Yang Etal[4] proposed a semi-supervised model of low-light image enhancement, which is enhanced in two stages. These methods are trained on paired data sets, and their enhancement performance largely depends on the data sets. Because the synthetic data can not fully describe the degradation in the real scene, and the real captured paired data contains a limited

© The 2022 International Conference on Artificial Life and Robotics (ICAROB2022), January 20 to 23, 2022

variety of scenes, the results of these methods are still imperfect, especially unable to deal with dense noise. Our model decomposes the complex image enhancement problem into sub-problem levels related to different features, which can be solved separately for multi-branch fusion, and the attention mechanism is embedded in the fusion process, which handling brightness/contrast enhancement and image denoising simultaneously.

3. Methodology

This paper proposes a novel end-to-end attention-guided method (A-MBLLEN) based on multi-branch convolutional neural network. The proposed network is composed with enhancement module (EM) and Convolutional Block Attention Module (CBAM). The attention module can make the CNN network structure gradually focus on the weak light area in the image, and the enhancement module can fully highlight the multi-branch feature graph under the guidance of attention. The overall network architecture and the data process flow is shown in Fig 1.

3.1. *Enhancement module (EM)*

EM contains multiple sub-nets, whose number equals to the number of the Branch numbers, and the output is a colorful image with the same size of the original low-light image. Each sub-net has a symmetric structure to

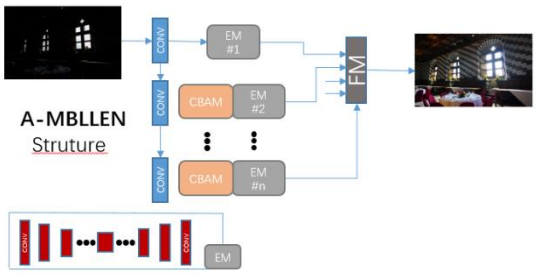

Fig 1. The workflow of the proposed approach for underexposed image restoration using A-MBLLEN. The proposed network with enhancement module (EM) and Convolutional Block Attention Module(CBAM) . The output image is produced via feature fusion.

apply convolutions and deconvolutions. The first convolutional layer uses 8 kernels of size 3×3, stride 1 and ReLU nonlinearity. Then, there are three convolutional layers and three deconvolutional layers, using kernel size 5×5, stride 1 and ReLU nonlinearity, with kernel numbers of 16, 16, 16, 16, 8 and 3 respectively. It should be noted that all subnets are trained at the same time, but are independent and do not share any learning parameters. The enhancement module can be considered to be composed of encoder and decoder. The encoder learns to extract the features of detection and weak light enhancement, and the decoder learns the accumulation from the feature space to the enhanced image. The encoder and decoder are connected skipply for detail reconstruction.

3.2. *CBAM*

Convolutional Block Attention Module (CBAM)[5] has two Attention submodules, CAM(Channel Attention Module) and SAM(Spatial Attention Module). CBAM details are shown below fig.2.

The CAM framework is shown in Fig3.

the input feature map F (H×W×C) is operated through global max pooling and global average pooling based on width and height, respectively, to obtain two 1×1×C feature maps which are separately feed into A two-layer multilayer perceptron (MLP). The number of neurons in

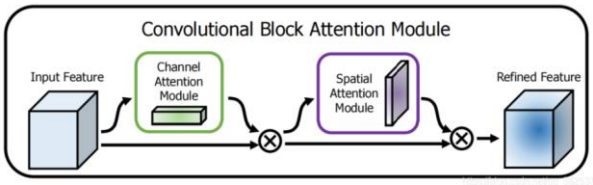

Fig 2. CAM(Channel Attention Module) and SAM(Spatial Attention Module). CAM is responsible for the attention weight on Channel, SAM is responsible for the attention weight on space (Height, Width).

the first layer is C/r (C is channel number, r is the

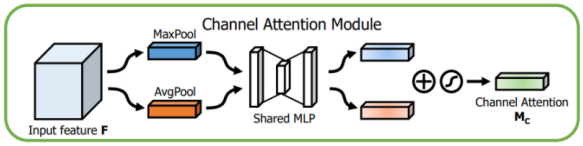

Fig 3. CAM struture

reduction rate), the activation function is ReLU, and the number of neurons in the second layer is C. The network is shared. Then, the MLP output features are subjected to an element-wise addition operation, and then the sigmoid

activation operation is performed to generate the final channel attention feature, namely $M_C(F)$ in formula(1).

$$M_C(F) = \sigma\left(W_1\left(W_0\left(F^C{}_{\text{avg}}\right)\right) + W_2\left(W_0\left(F^C{}_{max}\right)\right)\right) \quad (1)$$

where $F^C{}_{\text{avg}}$ is the result of global average pooling, and $F^C{}_{max}$ is the result of global maximum pooling.

The SAM framework is shown in Fig4.

The feature map F output by the channel attention module as the input to this module. First, do a channel-based global maximum pooling and global average pooling to obtain two feature maps of size H×W×1, and then perform connection operations on these two feature

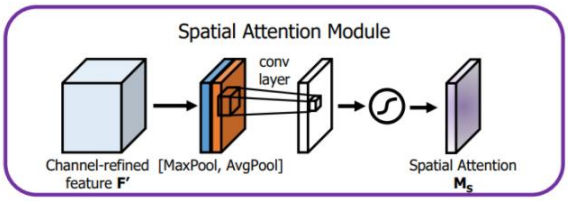

Fig 4 . SAM struture

maps based on channels. Then after a 7×7 convolution operation, the dimensionality is reduced to H×W×1, and then the spatial attention feature is generated through sigmoid, which is $M_s(F)$ in the formula (2).

$$M_s(F) = \sigma\left(f^{7\times7}[F^s{}_{\text{avg}}, F^s{}_{\max}]\right) \quad (2)$$

where f represents the convolution operation and [,] represents the channel concatenation operation. Finally, $M_s(F)$ is multiplied by the input feature map of the module to obtain the final generated feature map.

3.3. *Loss Function*

The function of loss function is to make the enhanced image $E_{(x,y)}$ of the input image $I_{(x,y)}$ after the trainable CNN with parameters W enhancement as close as possible to the input reference image $R_{(x,y)}$. To improve the image quality both qualitatively and quantitatively, we propose a novel loss function. The MSE loss function to be minimized as:

$$\text{MSE}=\frac{1}{n}\sum_i^n \left\| E^i{}_{(x,y)} - R^i{}_{(x,y)} \right\|_2 \quad (3)$$

The structural loss function adopts DSSIM which is derived from structural similarity (SSIM)[6] and can be expressed as:

$$\text{DSSIM}=\frac{1}{n}\sum_i^n (1 - SSIM(R^i{}_{(x,y)} - E^i{}_{(x,y)})) \quad (4)$$

The Context loss can improve the visual quality. Because the VGG network [7] is shown to be well-structured and well-behaved, we choose the VGG network as the content extractor in our method. the context loss is defined as follows:

$$L_{VGG} = \frac{1}{C_{i,\,j} H_{i,\,j} W_{i,\,j}} \sum_{x=1}^{C_{i,\,j}} \sum_{y=1}^{H_{i,\,j}} \sum_{z=1}^{W_{i,\,j}} \left\| \varphi_{i,j}(E)_{x,y,z} - \varphi_{i,j}(R)_{x,y,z} \right\| \quad (5)$$

where E and G are the enhanced image and ground truth, and $w_{\text{i,j}}$, $H_{\text{i,j}}$ and $C_{\text{i,j}}$ describe the dimensions of the respective feature maps within the VGG network. Besides, $\varphi_{i,j}$ indicates the feature map obtained by j-th convolution layer in i-th block in the VGG-19 Network

Total Loss. The total loss can be expressed as:

$$L_{total} = MSE + DSSIM + L_{VGG} \quad (6)$$

4. Experimental Evaluations

Our implementation is done with Keras (Chollet et al. 2015)and Tensorflow (Abadi et al. 2016). The proposed network can be quickly converged after being trained for 20 epochs on a Titan-XGPU using the proposed dataset. In the experiment, training is done using the ADAM optimizer [8] with a learning rate of α = 0.002, β_1 = 0.9, β_2 = 0.999 and $\varepsilon = 10^{-8}$.

4.1. *Dataset and Metrcis*

Having a large dataset with diverse scenes and lighting conditions is significant for training a well-generalized model. We select 16925 images in the VOC dataset to synthesize the training set, 56 images for the validation set, and 144 images for the test set, The synthesis method is from reference [9]. Our synthetic low-light data sets have advantages over real low-light data sets, as they may be subject to device and environment constraints with a wide variety of light levels. By using the reference images provided in our dataset as enhancement groundtruth, we are able to quantitatively evaluate different methods in terms of PSNR and SSIM [10] indices. As no-referenced image quality assessment, We adopt Average Brightness (AB) [11] and Natural Image Quality Evaluator (NIQE) [12] which is a well-known no-reference image quality assessment for evaluating real image restoration without ground-truth, to provide quantitative comparisons. We compare our method with other methods on our synthetic dataset, Quantitative comparison is shown in Table 1.

Table 1 Comparison of different models.

Models	PSNR ↑	SSIM ↑	AB ↑	NIQE ↑
RetinexNet[3]	23.66	0.747	-4.14	30.57
ElightenGAN[14]	24.07	0.827	-3.13	27.76
MBLLEN[15]	25.95	0.885	0.008	29.47
OURS	**26.57**	**0.894**	**0.010**	27.62

5. Conclusions

In Table 1, our method achieves almost all the best results under all quality indicators. The only case where it ranks

last is the brightness scaling result under the NIQE indicator. In general, our model has achieved good results from both qualitative and quantitative perspectives. We applied the model to actual noisy pictures, and found that A-MBLLEN can be adapted to real noisy low-light images to a certain extent, and can produce visually pleasing enhanced images. In the future, We will further explore better low-light adjustment methods to solve more complex realistic pictures.

References

1. A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Processing Magazine*, *vol. 35, no. 1*, 2018, pp. 53-65.
2. Kin Gwn Lore, Adedotun Akintayo, Soumik Sarkar, LLNet: A deep autoencoder approach to natural low-light image enhancement, *Pattern Recognition*, 2017, 61:pp.650-662.
3. Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep Retinex Decomposition for Low-Light Enhancement. In BMVC, 2018. 2, 5, 6, 7, 8
4. W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, "From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement," *CVPR*, 2020, pp. 3063–3072.
5. Woo, Sanghyun, et al. "Cbam: Convolutional block attention module." Proceedings of the European conference on computer vision (ECCV). 2018.
6. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, , 2004, pp. 600–612.
7. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
8. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
9. Lv F , Li Y , Lu F . Attention Guided Low-light Image Enhancement with a Large Scale Low-light Simulation Dataset[J]. 2019.
10. 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*, 2004,13(4) pp:600–612,.
11. ZhiYu Chen, Besma R Abidi, David L Page, and Mongi A Abidi. Gray-level grouping (glg): an automatic method for optimized image contrast enhancement-part i: the basic method. *IEEE transactions on image processing*, 2006, 15(8),pp :2290–2302,.
12. Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a" completely blind" image quality analyzer. *IEEE Signal Process*. 2013, Lett. 20(3) pp:209–212.
13. Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou,and Z. Wang, "EnlightenGAN: Deep light enhancement without paired supervision," 2019, arXiv arXiv:1906.06972.
14. F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: Low-light image/video enhancement using cnns," in BMVC, 2018.

Authors Introduction

Mr.Xiwen Liang

He received his bachelor's degree from the school of electronic information and automation of Tianjin University of science and technology in 2021. He is acquiring for his master's degree at Tianjin University of science and technology.

Mr. Xiaoning Yan

He graduated from Tianhua South University of Technology in 2012 with a bachelor of software engineering. In 2015, he received a master's degree in computer science from the University of Texas at Dallas. Since 2017, he has served as the technical director of Shenzhen Ansoft Vision Technology Co., Ltd.

Prof. Ms. Xiaoyan Chen

She, professor of Tianjin University of Science and Technology, graduated from Tianjin University with PH.D (2009), worked as a Post-doctor at Tianjin University (2009.5-2015.5). She had been in RPI, USA with Dr. Johnathon from Sep.2009 to Feb.2010 and in Kent, UK with Yong Yan from Sep-Dec.2012. She has researched electrical impedance tomography technology in monitoring lung ventilation for many years. Recently, her research team is focus on the novel methods through deep learning network models.

Mr. Hao Feng

He received his bachelor's degree from the school of electronic information and automation of Tianjin University of science and technology in 2020. He is acquiring for his master's degree at Tianjin University of science and technology.