SUGISAKA MASANORI

Journal of Robotics, Networking and Artificial Life
Vol. 9(4); March (2023), pp. 316–320
ISSN (Online):2352-6386; ISSN (Print):2405-9021
https://alife-robotics.org/jrnal.html

Research Article

Attention-guided Low light enhancement CNN Network

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

[1]*Tianjin University of Science and Technology, No. 1038 Dagu Nanlu, Hexi District, Tianjin, China, 300222*
[2]*Shenzhen softsz co. ltd, Shenzhen, China*

ARTICLE INFO

Article History
Received 25 November 2021
Accepted 19 October 2022

Keywords
Low-light image enhancement
Deep learning
Multi-branch fusion
Convolutional neural network

ABSTRACT

Low illumination image enhancement is a difficult but scientific task. With the image brightness increasing, the noises are amplified, and with the contrast and detail increasing, the false information is generated. To solve this problem, a multi-branch attention network is proposed to process low-light images directly without additional operations. 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 way, the overall quality of the picture will be greatly improved, including contrast, brightness, etc. Through a large number of experiments, our model can produce better visual effects, and also achieve good results in quantitative indicators.

1. Introduction

Due to unavoidable environments or technical limitations, many photographs are often taken under less than ideal lighting conditions. Poorly light 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 degradations 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. Since many existing algorithms basically brighten the whole picture, including traditional algorithms and deep learning algorithms, we consider introducing CBAM into the neural network to make the model have more self-attention. Extensive experiments show that the introduction of this attention mechanism can improve both visual perception and indicator estimation. Our contributions are summarized as follows. 1) We introduce the convolution block attention module (CBAM) into the multi branch enhancement module, which improve the feature extraction capability of the model without significantly increasing the amount of computation and parameters. 2) Our method is also effective in suppressing image noise and artifacts in low light region.

2. Related Work

Low-light image enhancement techniques have continued to evolve in recent decades. Low light is a low-quality image, and enhancing low-light images can not only improve the beauty of the image, but also help subsequent high-level visual tasks. Existing neural networks have some powerful tools such as end-to-end networks and GAN[1] to process low-light images.

Corresponding author's E-mail: cxywxr@tust.edu.cn, 1594838831@qq.com

LLNET[2] is the first real network to apply neural network to image enhancement. The paper uses deep self-encoding to extract image features and improve image brightness. Retinex-Net[3] splits the network into two modules, and achieves good results by pairing the constraints of the dataset and using the BM3D denoising module. In order to integrate the advantages of CNN and GAN, Yang Etal[4] proposed a semi-supervised model to enhance the image in two stages. Lv F Etal[15] proposed an attention-guided enhancement method and corresponding multi-branch network structure for simultaneous image enhancement and noise suppression. 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. Channel and spatial attention are embedded in the network, placed before each feature extraction module EM, which can improve the capability of the model in feature extraction without significantly increasing the amount of computation and parameters. Through this operation, brightness/contrast enhancement and image denoising are also better solved.

3. Methodology

This article proposes a hybrid attention mechanism framework A-MBLLEN based on multi-branch fusion. The proposed network is composed with enhancement module (EM) and Convolutional Block Attention

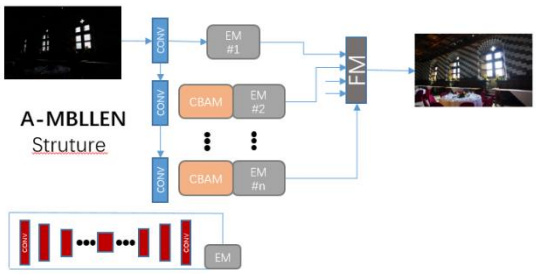

Fig 1. This is the proposed multi-branch attention network framework A-MBLLEN. The proposed network with enhancement module (EM) and Convolutional Block Attention Module (CBAM). After feature fusion, adjust the channel to 3 and output the result

Module (CBAM). The EM input is the output of the CBAM layer. U-Net[16] has been very successful in semantic segmentation, image restoration, and expansion. U-Net extracts multi-level features from different depth layers, retains a wealth of texture information, and uses multi-scale contextual information to synthesize high-quality images. I am using U-Net as the backbone of the EM network. 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 details of the entire framework are shown in Fig 1.

3.1. *Enhancement module (EM)*

EM contains multiple subnets, the number equals the number of branches, and the output color image is the same size as the input image. Each subnet has a symmetric structure for applying convolution and deconvolution. Their convolution kernel size is 3*3, the stride is 1, and the number of kernels is 8, 16, 16, 16, 16, 8, 3, respectively. the RELU nonlinear activation function is used to increase the model capacity. After passing through 9 EM modules, finally concatenate them together to get the fused features. 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 is considered composing of encoder and decoder. The first part is used to extract features in the image, and then skip the connection with the decoder to achieve information flow at the same scale. Finally, the information on multiple scales is fused to obtain an enhanced image. The encoder and decoder are skipping connected 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.

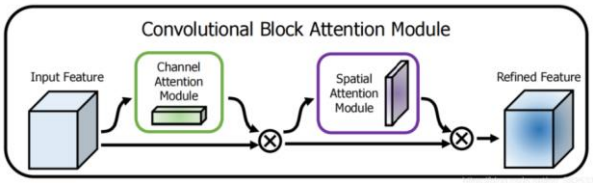

Fig 2. CAM(Channel Attention Module) and SAM(Spatial Attention Module). CAM is responsible for the attention weight on Channel, SAM is responsible for the attention weight on space (Height, Width).

The CAM framework is shown in Fig 3.

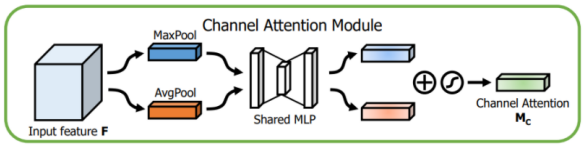

Fig 3. CAM struture

If the input image is H*W*C size, the input is subjected to global average pooling and global maximum pooling, respectively, to obtain a 1*1*C feature map, and then pass it into a two-layer perceptron. Channel attenuation is performed in the first layer to C/r, where c is the attenuation ratio. The second layer is still restored to the C channel, and its weights are shared. After two layers of perceptrons, through an element addition operation, and finally through the sigmoid activation function, the channel attention feature$M_C(F)$ is obtained.

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

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

The SAM framework is shown in Fig 4.

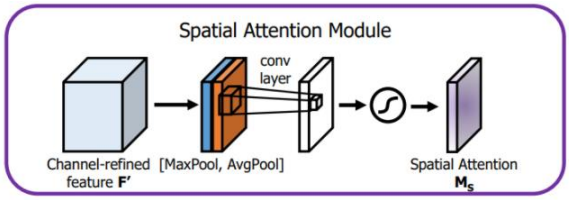

Fig 4. SAM struture

The output result F of the channel attention module is sent to the spatial attention module. After global maximum pooling and global average pooling, the two feature maps are connected in the channel direction to obtain the feature map of H*W*2. Finally, a 7*7 convolution kernel is used for channel dimension reduction, and finally a H*W*1 is output. Finally, the sigmoid activation function is performed to obtain the final spatial attention map M_s (F). The expression of M_s (F) is as Equation 2.

$$M_s(F) = \sigma\left(conv\left(concat\left(F^s{}_{\text{avg}}, F^s{}_{\max}\right)\right)\right) \quad (2)$$

where $concat(\,)$ denotes the concatenation operation; $conv(\,)$ denotes the convolution operation. Finally, $M_s(F)$ is multiplied by the input feature map of the module to obtain the final generated feature map.

By adding two dimensions of attention, it not only enhances the Structural similarity of the image, but also suppresses the noise.

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$R_{(x,y)}$. In order for the enhancement network to improve the image quality, we need a sophisticated loss function design to have a good effect in terms of quantity and quality. The MSE loss function to be minimized as:

$$\text{MSE}=\left\|E_{(x,y)} - R_{(x,y)}\right\|_2 \quad (3)$$

The structural loss function adopts DSSIM which is comes from Structural Similarity SSIM[6] and can be expressed as:

$$\text{DSSIM}= (1 - SSIM\left(R_{(x,y)} - E_{(x,y)}\right))/2 \quad (4)$$

We generally obtain the structural similarity of the pixel p in the two images by the following simple expression.

$$\text{SSIM}=-\frac{1}{N}\sum_{p\in img}\frac{2u_x u_y+C_1}{u_x{}^2+u_y{}^2+C_2}\cdot\frac{2\sigma_{xy}+C_2}{\sigma_x{}^2+\sigma_y{}^2+C_2} \quad (5)$$

where u_x and u_y are pixel value averages, $\sigma_x{}^2$ and$+\sigma_y{}^2$ are variances, σ_{xy} is covariance, and C1 and C2 are parameters to maintain the stability of SSIM.

The Context loss can improve the visual quality. We choose a VGG network[7] with stable structure to ensure that the extracted features are more representative of the original image. the context loss is defined as follows:

$$L_{VGG} = \frac{1}{C_{i,\,j}H_{i,\,j}W_{i,\,j}}\sum_{x=1}^{C_{i,\,j}}\sum_{y=1}^{H_{i,\,j}}\sum_{z=1}^{W_{i,\,j}}\left\|\varphi_{i,j}(E)_{x,y,z} - \varphi_{i,j}(R)_{x,y,z}\right\| \quad (6)$$

where E and G represent the enhanced image and the reference image, respectively. $w_{\text{i,j}}$, $H_{\text{i,j}}$ and $C_{\text{i,j}}$ is the feature map under different dimensions in VGG[7]. Besides, $\varphi_{i,j}$ represents the result output from the first convolutions in several blocks in the VGG-19 Network. The total loss can be expressed as:

$$L_{total} = MSE + DSSIM + L_{VGG} \quad (7)$$

4. Experimental Evaluations

We are implementing on tensorflow and Keras. Using pretrained weights we can fit faster. During training, we use ADAM[8] optimizer and set parameters a = 0.002, $\varepsilon = 10^{-8}$, β_1 = 0.9 and β_2 = 0.999.

4.1. *Dataset and Metrcis*

It is extremely important for the model to have a training set that is rich in variety. We pick 16,295 images from the VOC[17] dataset and use a scientific approach[9] to synthesize low-light and noisy paired images. With reference images in our dataset, we are able to perform qualitative and quantitative enhancement assessments. Among the evaluation methods with reference, we choose SSIM[10] and PSNR, and the evaluation indicators without reference include brightness average AB[11] and natural image quality evaluation NIQE[12]. These metrics are good for an objective evaluation of the enhancement effect, and we quantitatively compare the model(EnlightenGAN[13], MBLLEN[14]) proposed in this paper with several existing models. The results are 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
EnlightenGAN	24.07	0.827	-3.13	27.76
MBLLEN	25.95	0.885	0.008	29.47
OURS	**26.57**	**0.894**	**0.010**	27.62

Image Structural Similarity (SSIM) as a measure of Similarity between two images by calculating the SSIM of the enhanced results of the reference image and low illuminance image, the performance of the algorithm can be analyzed. Generally speaking, the larger SSIM value is, the closer it is to the reference image, that is, the illumination enhanced the better the results. After the algorithm proposed in this paper enhanced the low-illumination image, SSIM value was improved to 0.894, which is superior to other methods, indicating that the model proposed in this paper can improve the structural similarity between the enhanced image and the real image. Peak Signal-to-noise Ratio (PSNR) is a commonly used index to evaluate image quality. It reflects the degree of distortion between the image restored by low illumination image algorithm and the reference image. Generally speaking, The larger the PSNR, the better the picture fidelity, that is, the less the image is distorted compared to the reference image. As can be seen from the table, after using the image enhancement algorithm proposed in this paper, The PSNR is improved to 26.57, which has obvious advantages compared with other algorithms, thus proving the effectiveness of our proposed algorithm.

5. Conclusions

As shown in Table 1, we basically have the best results on all indicators, and one of our indicators, NIQE, is ranked last. 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 model noisy pictures in real environments, resulting in enhanced images with good visual perception. In the future, we will develop more useful algorithms to solve real-world low-light image enhancement problems, such as Multi-scale fusion of low illumination image features and some segmentation and noise image as prior information enhance the enhancement and generalization ability of the model.

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. cF. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: Low-light image/video enhancement using cnns," in BMVC, 2018.
15. Lv F, Li Y, Lu F. Attention Guided Low-Light Image Enhancement with a Large Scale Low-Light Simulation Dataset[J]. International Journal of Computer Vision, 2021, 129(11).
16. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. Unet: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted in1tervention, pages 234–241. Springer, 2015. 4.
17. Mark Everingham, Luc V an Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. International journal of computer vision, 88(2):303–338, 2010

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. Nenghua Xu

He, general Manager of Shenzhen Ansoft Technology Co. , Ltd., graduated from South China University of Technology, expert in Huawei's range of engineering computing acceleration and artificial intelligence. He has worked as a senior engineer in Huawei and a high-level innovative entrepreneur in Shenzhen.

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.