An Efficient Extreme-Exposure Image Fusion Method

Jiebin Zhang\textsuperscript{1,a}, Shangyou Zeng\textsuperscript{1,*}, Ying Wang\textsuperscript{1}, Jinjin Wang\textsuperscript{1}, Hongyang Chen\textsuperscript{1}

\textsuperscript{1}School of electronic engineering, Guangxi Normal University, Guilin, Guangxi, 541001, China

\textsuperscript{a}email: 13406248995@163.com

\textsuperscript{*}Corresponding author’s e-mail: zsy@mailbox.gxnu.edu.cn

Abstract. Since the existing commercial imaging equipment cannot meet the requirements of high dynamic range, multi-exposure image fusion is an economical and fast method to implement HDR. However, the existing multi-exposure image fusion algorithms have the problems of long fusion time and large data storage. We propose an extreme exposure image fusion method based on deep learning. In this method, two extreme exposure image sequences are sent to the network, channel and spatial attention mechanisms are introduced to automatically learn and optimize the weights, and the optimal fusion weights are output. In addition, the model in this paper adopts real-value training and makes the output closer to the real value through a new custom loss function. Experimental results show that this method is superior to existing methods in both objective and subjective aspects.

1. Introduction

Dynamic range (DR) is called the ratio of maximum light intensity to minimum light intensity, which is $2^{24}$ in natural scenes. However, the current commercial digital SLR camera (DSLR) can only capture scenes with a low dynamic range of $2^{8} - 2^{12}$. In other words, it is very difficult to record a good high dynamic range (HDR) with current imaging devices. This disadvantage will lead to underexposed or overexposed areas in the captured image. Therefore, in order to better record objects with different brightness, multi exposure image fusion (MEF) is a fast and economical way in the field of HDR imaging [1].

MEF algorithm is suitable for multiple image sequences with different exposure levels of the same scene. An excellent MEF algorithm is expected to capture some information from different exposure images and fuse these important perceptual information into an HDR image, which has the best visual effect and the best objective evaluation index[2].

Burt et al[3] proposed the classical fusion algorithm of image pyramid, which calculates the weight according to the correlation between local energy and pyramid. Goshtasby et al[4] blocked the exposed image and select each block with the highest information. However, this is easily affected by block defects. At the same time, with the explosion of deep learning, more and more people use deep learning methods to study multi exposure image fusion tasks like Prabhakaret al[5] first proposed an unsupervised MEF learning framework, which uses SSIM (structural similarity) loss function to measure image quality. These methods greatly promote the development of multi exposure image fusion task, and provide method guidance for the extreme exposure image fusion method proposed in this paper.

The main work of this paper is as follows:

(i). A new extreme-exposure image fusion model is designed, which greatly reduces the storage space and running time.
(ii). customized loss function to improve network performance.
(iii). The SimAM attention module is introduced to weight the deep features of the encoder.

The rest of this article is organized as follows. The second part introduces the fusion model of the classical method and the convolutional neural network (CNN). The third part is the framework of extreme exposure image fusion model. The fourth and fifth parts are the experimental results and summary respectively.

2. Related Work
The purpose of multi exposure image fusion is to generate well exposed images $y_f(c)$ from a group of images $X_i(c)$ with different exposure levels, which can be expressed as:

$$y_f(c) = \sum_{i=1}^{l} \omega_i(c) \ast X_i(c)$$

Where $\omega_i(c)$ is the fused weight and $c \in \{R,G,B\}$ represents the channel of color images. Different MEF algorithms are different in weight calculation.

In recent years, deep learning has made a great leap in various machine vision tasks like image classification [6], face recognition [7], semantic segmentation [8], object detection [9] and much more. Liet al [10] proposed a new Densefuse architecture, which fuses HDR image sequences for infrared and visible images. Ma al [11] proposed a fast MEF network method for static image sequences with arbitrary spatial resolution and exposure times. This method performs end-to-end training by optimizing the perceptual calibrated MEF structure similarity (MEF-SSIM) index on the database. In addition, a pilot filter is used to improve the performance. Zhang al [12] proposed a general image fusion framework ifcnn based on convolutional neural network. Inspired by the transform domain image fusion algorithm, two convolution layers are used to extract significant image features from multiple input images. Then, different fusion rules (elementwise max, elementwise min or elementwise mean) are selected according to the type of input image to fuse the convolution features of multiple input images.

Although deep learning based methods have been proposed, and their performance is better than traditional methods, these methods often require multiple image sequences with different exposure levels, which greatly increases the memory space and runtime, and the fusion effect is not ideal when there are only two extreme exposure images.

![Figure 1. Architecture of the proposed network.](image)

3. Proposed Method
In order to solve these problems, we propose an extreme exposure image fusion method. Only two extreme exposure images can output the best fusion weight and obtain a well exposed fusion image, which is more suitable for practical application.
3.1. Network Architecture
The model of this work is shown in Fig. 1. Firstly, two extremely exposed images are input into the network, the RGB three channels are separated and reorganized, and then input into the encoder densenet block module to fully extract the pixel features of the image. Then, the perceived importance of the features input to the simam attention module is weighted and sent to the decoding module network for further extraction and reconstruction, so as to obtain a well exposed HDR image.

3.2. Loss Function
In this model, we propose a new custom loss function for extreme exposure image fusion task to improve performance. The formula of multi-scale weighted loss function is as follows:

\[
\text{loss} = \text{loss}_{\text{MSE}} + 1000 \times (1 - \gamma) \times \text{loss}_{\text{SSIM}}
\]

\[
\text{loss}_{\text{SSIM}}(x, y; \sigma) = -\log \left( \frac{\left( \frac{2\sigma_x \sigma_y + C_1}{\sigma_x^2 + \sigma_y^2 + C_1} \right) \left( \frac{2\sigma_x \sigma_y + C_1}{\sigma_x^2 + \sigma_y^2 + C_1} \right)}{\left( \frac{1}{2m} \sum_{i=1}^{m} (x_i - y_i)^2 \right)^2} \right)
\]

\[
\text{loss}_{\text{MSE}}(x, y) = \frac{1}{2m} \sum_{i=1}^{m} (x_i - y_i)^2
\]

Where \( C_1, C_2 \) are constants, and the experiment set as \( 1 \times 10^{-4} \) and \( 9 \times 10^{-4} \) respectively. \( \sigma_x \) is the region of image \( x \) within window \( \sigma_x \), and \( \bar{\sigma}_x \) is the mean of \( \sigma_x \). The variables \( \sigma_x^2 \) and \( \sigma_y^2 \) are the variance of \( \sigma_x \), and the con-variance of \( \sigma_x \) and \( \sigma_y \). Similarly, \( \bar{\sigma}_y \), \( \sigma_y^2 \), and \( \sigma_{xy}^2 \) represent the corresponding meaning. \( x_i \) and \( y_i \) represent the predicted value and real value. \( \gamma \) is the weight coefficient. Thus, \( \gamma \) can be defined as follows:

\[
\gamma = \frac{\eta(\sigma_{xy}^2)}{\eta(\sigma_{x}^2) + \eta(\sigma_{xy}^2)}
\]

Where \( \eta(x) = \max(x, 0.0001) \) is a selection correction function used to increase the stability of the network structure.

3.3. SimAM Attention
We add a simple but effective convolutional neural network attention module, as shown in Figure 2, which is different from the existing channel and spatial attention module CBAM [13]. Without adding parameters to the original network, the module infers three-dimensional attention weights for the feature graphics in the layer, which improves the attention accuracy[14].

![Figure 2. Full 3-D weights for attention.](image)
4. Experimental Details
In this part, we first introduce the details of data processing method and network training, which is a necessary preparation for the experiment. Secondly, based on the same image input of two extreme exposures, the model is compared with the classical and latest MEF methods in terms of visual quality and objective indicators.

4.1. Preparation
In this work, two extreme exposure images are selected from the SICE[15] common data set with images of different exposure levels to form the SICE-ex data set. In this paper, Adam optimizer is selected, the epoch is set to 12, the batch is 4, and the initial value of learning rate is 1e-4. The learning rate is corrected by 50% or 105% according to the training cycle.

![Figure 3. Three groups of unreferenced exposure image sequences.](image)

![Figure 4. Compared with the classical algorithm Mentenrs09 and the method based on deep learning, the test results of the proposed method in the input extreme exposure image sequence show that the exposure is good and the image is clear.](image)

![Table 1. The method proposed in this paper is based on the numerical analysis of four target indexes, classical algorithm menter09 [16] and the most advanced MEF algorithms li19 [17], yang18 [18], lee18 [19] and li20 [20]. Bold is the best result](image)

|                | Our  | Li19 | Yang18 | Merterns09 | Lee18 | Li20 |
|----------------|------|------|--------|------------|-------|------|
| NIQE averge    | 3.0019 | 3.4395 | 3.1517 | 3.1717     | 3.1667 | 3.2201 |
| Time averge    | 0.2521 | 1.8116 | 0.2762 | 0.8142     | 0.3574 | /    |
| figure gradient| 9.0609 | 5.3362 | 8.8521 | 8.5242     | 8.0599 | 8.8425 |
| ave gradient   | 9.1476 | 5.2919 | 8.7732 | 8.4979     | 8.0136 | 8.7301 |

4.2. Experimental Result
This paper randomly tests the non reference polarity exposure image sequence, and compares it with menter09 [16], li19 [17], Yang18 [18], Lee18 [19], li20 [20] from subjective vision (Fig. 3 and Fig. 4) to fusion duration, average gradient and graphic gradient, and Natural Image Quality Evaluator[21] as follow Table 1.
5. Conclusion
This paper proposes an extreme exposed image fusion network which is completely different from MEF fusion algorithm. The performance of the network is improved by introducing spatial and channel double attention mechanism and user-defined loss function. On the premise of ensuring clear image structure and good exposure, the image storage space and fusion time are greatly saved, and the embeddability of the algorithm on mobile terminals is improved. A large number of experiments show that the fusion method is superior to existing methods in subjective vision and objective evaluation. However, due to the use of two extreme exposure images, the color, depth of field and contrast of fusion image are slightly inadequate compared with other multi-exposure image fusion algorithms, so the improvement of these directions will get more attention in the future.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (61976063).

References
[1] Reinhard,E., Heidrich,W., Debevec,P. (2010)High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting. Morgan Kaufmann. San Francisco.
[2] Ma,K., Zeng,K., and Wang,Z. (2015)Perceptual quality assessment for multi-exposure image fusion. In: IEEE Trans.Image Process. pp. 3345–3356.
[3] Burt,P.J., and Koczynski,R.J. (1993)Enhanced image capture through fusion. In: International Conference on Computer Vision. pp. 173–182.
[4] Goshtasby,A. (2005)Fusion of multi-exposure images. In: Image and Vision Computing. pp. 611–618.
[5] Prabhakar,K.R., Srikar,V.S., and Babu,R.V. (2017)DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In: International Conference on Computer Vision. pp. 4724–4732.
[6] Kang, X.D, Cheng,C. (2018)Classification of Hyperspectral Images by Gabor Filtering Based Deep Network. In: IEEE Journal of Selected Topics in Applied Earth. pp. 1166-1178.
[7] Chan,T.H., Jia,K., Gao,S.H., (2015)PCANet: A Simple Deep Learning Baseline for Image Classification?. In: IEEE Trans.Image Process. pp. 5017 - 5032.
[8] Akkus, Z., Galimzianova,A., Hoogi, A. (2017)Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions. In: Journal of Digital Imaging. pp. 449-459.
[9] Zhao,Z.Q., Zheng,P. (2019)Object Detection With Deep Learning: A Review. In: IEEE Transactions on Neural Networks and Learning Systems. pp. 3212-3232.
[10] Li,H., and Wu,X. (2019)DenseFuse: A fusion approach to infrared and visible images. In: IEEE Trans.Image Process. pp. 2614–2623.
[11] Ma,K.D.,Duanm,Z.D.,Zhu,H.W.(2020) Deep Guided Learning for Fast Multi-Exposure Image Fusion. In: IEEE Trans.Image Process. pp. 2808–2819.
[12] Zhang,Y.,Liu,Y.,Sun,P.,Yan,H.(2020) IFCNN: A general image fusion framework based on convolutional neural network. In: Information Fusion. pp. 99–118.
[13] Cui,Z.Y., Li,Q., Cao,Z.J., Liu,N.Y. (2019)Dense Attention Pyramid Networks for Multi-Scale Ship Detection in SAR Images. In: IEEE Transactions on Geoscience and Remote Sensing. pp. 8983-8997.
[14] Yang,L.X., Zhang,R.Y. (2021)SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In: International Conference on Machine Learning. pp. 11863—11874.
[15] Cai,J.R., Gu,S.H., Zhang,L. (2019)Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. In: IEEE Trans. Image Process. pp. 2049-2062.
[16] Mertens,T., Kautz,J. (2001)Exposure Fusion: A Simple and Practical Alternative to High Dynamic Range Photography. In: Computer Graphics Forum. pp. 161-171.
[17] Li, H., and Wu, X.J. (2019) DenseFuse: A Fusion Approach to Infrared and Visible Images. In: IEEE Trans. Image Process. pp. 2614-2623.

[18] Yang, Y., Cao, W., Wu, S.Q., and Li, Z.G. (2018) Multi-Scale Fusion of Two Large-Exposure-Ratio Images. In: IEEE Signal Processing Letters. pp. 1885-1889.

[19] Lee, S.H., Park, J.S., Cho, N.I. (2018) A Multi-Exposure Image Fusion Based on the Adaptive Weights Reflecting the Relative Pixel Intensity and Global Gradient. In: IEEE International Conference on Image Processing. pp. 1737-1741.

[20] Li, H., Ma, K., Yong, H.W. (2020) Fast Multi-Scale Structural Patch Decomposition for Multi-Exposure Image Fusion. In: IEEE Trans. Image Process. pp. 5805-5816.

[21] Mittal, A., Soundararajan, R., and Bovik, A.C. (2013) Making a “Completely Blind” Image Quality Analyzer. In: IEEE Signal Processing Letters. pp. 209-212.