IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-2010-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IGARSS52108.2023.10281494

HYBRID DENSELY CONNECTED NETWORK FOR MULTI-EXPOSURE IMAGE FUSION

Hao Zeng[1], *Yulei Wang*[1], *Haoyang Yu*[1], *Meiping Song*[1], *Enyu Zhao*[1], *and Tingting Tao*[2]

[1] Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian 116026, China
[2] Environmental Sciences and Engineering College, Dalian Maritime University, Dalian 116026, China

ABSTRACT

Multi-exposure image fusion (MEF) technique is the most widely used method to obtain high dynamic range (HDR) images. Inspired by the recent successful application of Transformer in image processing, a hybrid dense connection network based on CNN and Transformer is proposed for MEF in this paper. Considering the importance of texture details to the multi-exposure image fusion task, shallow features containing rich texture details is also added to each dense layer, which are extracted by the pre-trained RepVGG. In addition, the dynamic weight calculation module is improved, so that different source images can obtain finer weight in the calculation of the loss function. Experiments are conducted on the dataset provided by MEFB, and both qualitative and quantitative comparisons show that the proposed method can achieve better results compared with the state-of-the-art algorithms.

Index Terms— Multi-exposure image fusion, Transformer, unsupervised learning, hybrid dense connection.

1. INTRODUCTION

The multi-exposure fusion (MEF) technique can extract useful information from a set of low dynamic range (LDR) images with different exposures and fuse them into a high dynamic range (HDR) image, which provides a cost-effective method for obtaining HDR images, and has become the most widely used HDR image acquisition solution.

In the earliest studies, some conventional algorithms have been proposed, which can be divided into two categories, spatial domain-based and transform-domain based algorithms. Recently, with the development of modern techniques for deep learning, the convolutional neural network based MEF has attracted more attention. Xu et al. [1] proposed a FusionDN algorithm, which is a densely connected network that unifies various fusion tasks by applying elastic weights. The same group proposed U2Fusion [2] by improving the information preservation degree allocation strategy and loss function. The improvement of U2Fusion is that the degree of information preservation is determined by the information measurement of the extracted features by the pre-trained network instead of the amount and quality of the information in source images. However, these traditional CNN-based methods have a small receptive field with weak ability to extract global information, which easily leads to too smooth and dark changes in the brightness of the fused image and does not match the real scenario. To solve the above problem, the Transformer network with powerful global information extraction capability has been introduced into the MEF task. For example, Qu et al [3] proposed TransMEF, which combines VIT and CNN into an encoder to encode a single input, fuses multiple encoder outputs by a mean fusion strategy, and finally decodes the fused image by a decoder. However, due to the characteristics of VIT, TransMEF can only fuse small square images with integer multiples of 16 in length and width, which greatly limits its application. Furthermore, the computational complexity of the general Transformer increases quadratically with the increase of spatial resolution, so it cannot be well applied to the MEF task with high-resolution images.

In this paper, a new hybrid densely connected network is proposed to address the problems in the traditional CNN-based and the recently proposed Transformer-based MEF algorithms. Restormer, proposed by Zamir et al. [4] reduces the computational complexity by optimizing the multiheaded attention block (MSA) and the feedforward networks (FFN), making it possible to process high-resolution images. Therefore, it is introduced into the MEF task in this paper. First, the Restormer is cascaded with the CNN, and then additionally connected shallow features containing rich texture details output by the pre-trained RepVGG [5]. This new densely connected layer with different size perceptual fields ensures that local information is obtained without losing global information, making the global brightness variation of the fusion result more reasonable, while the connection of additional shallow features allows for clearer edge details. Finally, fine-grained loss functions and dynamic weight assignment module are designed for unsupervised learning.

2. METHODOLOGY

As shown in Fig.1, a hybrid densely connected network is designed for MEF task. It consists of three main components, the first one is a densely connected backbone network responsible for source image sequence feature extraction; the second one is a dynamic weight assignment module, whose role is to determine the weight of each source image in the calculation of the loss function; and finally, an image reconstruction module responsible for the inductive reconstruction of the extracted features. This section describes in detail the composition of the network structure parameters and loss function designed in this paper.

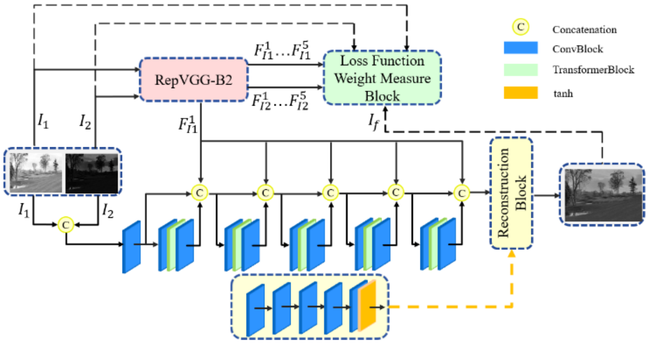

Fig.1. The framework of the proposed method.

2.1. Network design

Since the advent of DenseNet [6], it has been widely used by virtue of fewer parameters, better performance, and feature reuse by splicing. As a result, the architecture of this network is used as the main model in this paper and densely connected network is constructed for MEF tasks.

Each densely connected layer of DenseNet in this paper consists of two convolutional blocks and one Transformer block. The convolutional block consists of a convolutional layer, a batch normalization layer, and a LeakyReLu activation function. The size of kernel function in the convolutional layer is set to 3×3, and the step size is 1. The parameters of the two convolutional blocks in each densely connected layer are the same. The Transformer block consists of a convolutional layer together with the Restormer. In order to solve the problem of huge computation amount of Transformer, a 1×1 convolutional layer is designed in front of Restormer to make the channel of the feature map reduce to an acceptable size. In addition, since 5 modules containing Restormer have been cascaded with good global information extraction ability in both shallow features and deep features, the number of Transformer blocks per scale within a single Restormer is reduced to 2/3 of the original number, and the number of self-attended heads is reduced to 1/2 of the original number.

Since the pooling layers will lose most of the detailed information of the feature map, which is intolerable for MEF, the two densely connected layers are connected directly instead of using a transition layer containing the maximum pooling layer between every two densely connected layers. In addition, the Restormer used in this paper is much more capable of integrating information than shallow convolution, shallow texture features from the source image is stitched at the input of each densely connected layer, so that the local features such as texture details can be retained from being diluted, where the shallow texture features are extracted by the first "VGG block" of RepVGG-B2. Although the underexposed image contains more texture details, 1/4 of the stitched shallow texture features are from the overexposed image in order to avoid the low brightness of the fused image after adding the shallow texture features.

The image reconstruction module consists of five convolutional blocks with the same parameters except for the number of input and output channels. The size of the kernel function of the convolutional layer is set to 3×3 with a step size of 1. The activation function of the last convolutional block is replaced by the tanh activation function by the LeakyReLu.

For the unsupervised algorithm of the MEF task, the loss function is the sum of several different types of difference values between the fused image and the source image sequence, and the weights of each source image are the same. However, it is obvious that the levels of interest are different for different source graphics. Inspired by U2Fusion [2], the weight assignment strategy has been modified based on the measure of information retention, making it more suitable for multi-exposure fusion tasks specifically.

2.2. Loss function

The multi-exposure image fusion problem has been generalized as maintaining the structural similarity, gradient, and intensity between the source image and the fused image. Based on this, a loss function including these three terms have been designed in this paper.

The structural similarity (SSIM) metric reflects the distortion degree between the fused image and the source image in three aspects: luminance, contrast, and structure. Considering that in the MEF task, the luminance of the fused image is between multiple source images, the calculation of luminance will make the SSIM value large, which is not conducive for model convergence, and the larger luminance component makes the model pay less attention to the other two components. As a result, the luminance component has been removed from the SSIM. The other two items are the differential gradient and intensity between the fused image and the source image. The weights of the three components are set to 10, 25, and 25, respectively.

3. EXPERIMENTS

3.1. Training and Testing

Our network is trained on the SICE dataset, using the Y channel only with the image size cropped to 256 × 256, and the batch size is set to 8. The initial learning rate is set to 0.0001, and the learning rate is adjusted using a cosine annealing strategy.

The proposed network in tested on the MEFB benchmark test set provided by Zhang [7]. Due to the down-sampling step of Restormer, the input image size must be able to be divisible by 8, and as a result, reflection padding is performed on the upper and right edges of the source image sequence.

Finally, the fused image is post-processed by cropping off the new row and column pixels (whose values are 1-7) at the edges to meet the input size requirement, which is easy to implement. The fill operation is very common in convolutional neural networks, which is believed to be a negligible impact on the fusion results.

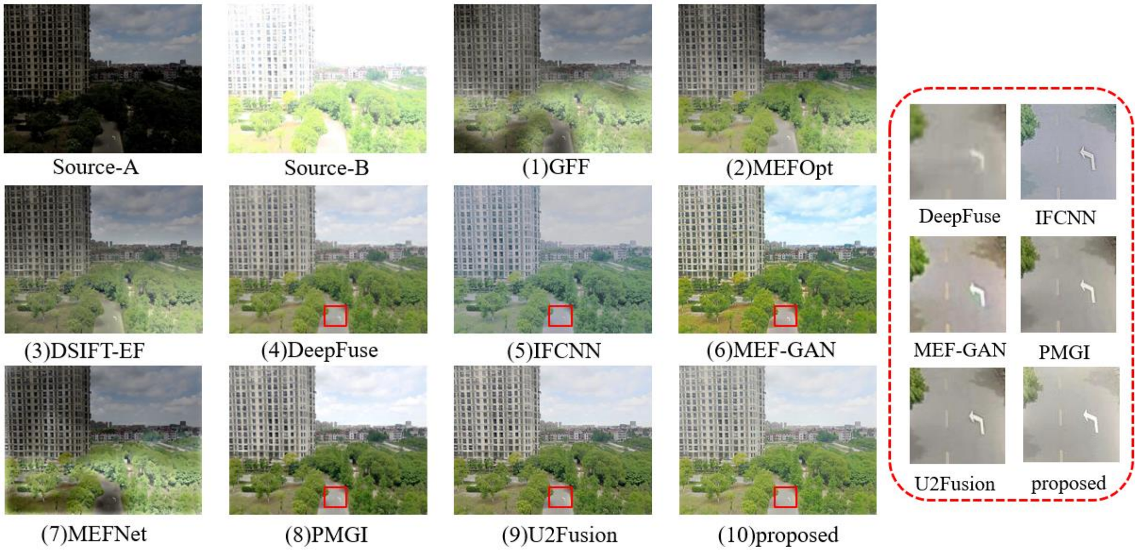

Fig.2. The fusion results of our algorithm and the comparison algorithm on the buildingRoad image pair. source-A is the underexposed image and source-B is the overexposed image. (1)-(3) are conventional algorithms and (4)-(10) are deep learning based algorithms. The red dashed box on the right side is a zoomed-in view of the red rectangular area on the left side.

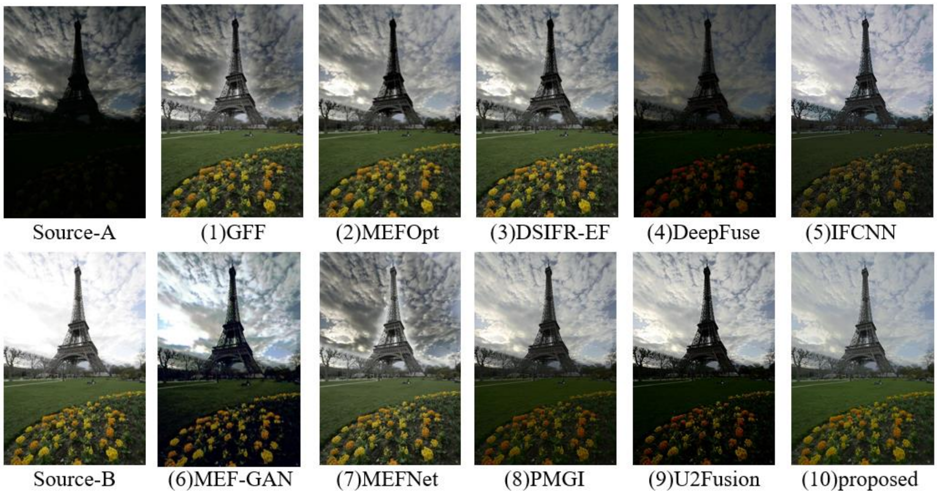

Fig.3. The fusion results of our algorithm and the comparison algorithm on the tower image pair.

3.2. Qualitative performance comparison

Fig.2 lists the fusion results of the proposed algorithm with some state-of-the-art algorithms on an outdoor buildingRoad image pair. The fusion results of the three conventional algorithms and deep learning-based MEFNet [8] perform poorly, and their adjacent local regions have too abrupt luminance changes and poor overall visual performance. The overall view of DeepFuse seems to be good, but the enlarged local view of arrows at the red rectangle can be seen to be with blurred details, which is the same situation as MEF-GAN [9] (even though its fusion results have the most vivid tones). IFCNN [10] has clear details in the arrows, but its color distortion is severe and some local areas are blurred (such as the sky). The overall visual performance of PMGI [11] and U2Fusion [2] performs well, but the images fused by the proposed method are clearer in the details. As can be seen in Fig.3, even though PMGI and U2Fusion perform well on buildingRoad image pair in Fig.2, they perform poorly on the other image in Fig.3, whose overall scene brightness is too dark with a poor visual perception. In conclusion, in terms of the results of the two qualitative experiments, the proposed model is more robust with higher generalization.

3.3. Quantitative performance comparison

Table.1 lists the performance of our algorithm and other comparative algorithms on 6 metrics, where the optimal values are marked in bold and the suboptimal values are marked with underlines. It is obviously seen that, the proposed algorithm can achieve best performance on the first five metrics, with only the last metric Q_{NCIE} to be the suboptimal value, where the gap between the suboptimal value and the optimal value is much smaller than the difference between the optimal value and the optimal value gaps of other algorithms.

Table.1. Quantitative evaluation of each algorithm on the buildingRoad image pair.

	Method	MEF-SSIM	PSNR	Q_{CV} ↓	NMI	$Q^{AB/F}$	Q_{NCIE}
Conventional	GFF	0.93543	57.283	781.03	0.17399	0.46125	0.80324
	MEFopt	0.85161	57.323	709.77	0.22442	0.45899	0.80375
	DSIFR-EF	0.89208	57.141	580.96	0.23390	0.43708	0.80396
Deep Learning-based	DeepFuse	0.87018	57.795	288.78	0.45943	0.17740	0.80676
	IFCNN	0.93520	57.840	146.44	0.47868	0.45735	0.80703
	MEF-GAN	0.67372	57.610	489.15	0.36810	0.20954	0.80549
	MEFNet	0.90312	57.284	558.48	0.21545	0.47471	0.80374
	PMGI	0.93499	57.621	520.63	0.54321	0.44918	**0.80881**
	U2Fusion	0.91736	57.722	351.41	0.49611	0.41974	0.80776
	proposed	**0.93624**	**57.850**	**128.15**	**0.55363**	**0.48168**	0.80853

4. CONCLUSION

To solve various problems in the current CNN-based and Transformer-based MEF algorithms, a new hybrid densely connected network has been proposed in this paper. The network is able to simultaneously acquire and fuse local feature information with global information and employs dense connectivity to improve the reuse of the extracted features while improving the information retention measure to make the dynamic weight assignment strategy applicable to a separate MEF task. Both qualitative and quantitative experiments on the MEFB dataset show that our approach outperforms the current state-of-the-art MEF algorithms.

5. ACKNOWLEDGEMENTS

This work is supported by the National Nature Science Foundation of China (61801075, 42101350, 42271355), the Natural Science Foundation of Liaoning Province (2022-MS-160), the China Postdoctoral Science Foundation (No. 2020M670723), and the Fundamental Research Funds for the Central Universities (3132023238).

6. REFERENCES

[1] Xu, Han, et al. "Fusiondn: A unified densely connected network for image fusion." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. No. 07. 2020.

[2] Xu, Han, et al. "U2Fusion: A unified unsupervised image fusion network." IEEE Transactions on Pattern Analysis and Machine Intelligence 44.1 (2020): 502-518.

[3] Qu, Linhao, et al. "Transmef: A transformer-based multi-exposure image fusion framework using self-supervised multi-task learning." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36. No. 2. 2022.

[4] Zamir, Syed Waqas, et al. "Restormer: Efficient transformer for high-resolution image restoration." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[5] Ding, Xiaohan, et al. "Repvgg: Making vgg-style convnets great again." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[6] Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[7] Zhang, Xingchen. "Benchmarking and comparing multi-exposure image fusion algorithms." Information Fusion 74 (2021): 111-131.

[8] Ma, Kede, et al. "Deep guided learning for fast multi-exposure image fusion." IEEE Transactions on Image Processing 29 (2019): 2808-2819.

[9] Xu, Han, Jiayi Ma, and Xiao-Ping Zhang. "MEF-GAN: Multi-exposure image fusion via generative adversarial networks." IEEE Transactions on Image Processing 29 (2020): 7203-7216.

[10] Zhang, Yu, et al. "IFCNN: A general image fusion framework based on convolutional neural network." Information Fusion 54 (2020): 99-118.

[11] Zhang, Hao, et al. "Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. No. 07. 2020.