Coupled Feedback Attention Networks

Rong Wang[1*], Chunjiang Duanmu[2]

[1]*College of Mathematics and Computer Science, Zhejiang Normal University, Jin Hua, Zhejiang, China*
[2]*College of Physics and Electronic Information Engineering, Zhejiang Normal University, Jin Hua, Zhejiang, China*

Abstract

In their daily lives, people frequently need to obtain images with a high dynamic range and resolution. Due to technological equipment limitations, high dynamic range images are produced by multi-exposure fusion (MEF) of low dynamic range images, while high resolution images are frequently obtained by super-resolution (SR) of low resolution images. MEF and SR are often analyzed separately. This research examines existing approaches and proposes a coupled feedback network attention network and its method to address the issue that current models cannot achieve high dynamic range and high resolution simultaneously.

Keywords

channel attention mechanism; coupled feedback mechanism;

1 Introduction

High dynamic range (HDR) images contain a broader dynamic range and richer texture features compared to typical low dynamic range (LDR) images and low resolution (LR) images, and high resolution (HR) images can enhance object detection accuracy. Technical methods to obtain HDR images and HR images, respectively, include single image super resolution (SISR) and multi-exposure image fusion (MEF).

By fusing two LDR images, the extreme exposure image fusion method creates an HDR image. Ma et al.[8] provided a quick approach for fusing multiple exposure images that improved the initial weights using a guided filter. Later, Xu et al[7] proposed a unified unsupervised fusion method that overcomes the fusion barrier of most images by constraining the similarity between the fused image and the original image.

With the continuous development of deep neural networks, many CNN-based methods have been proposed in the field of SISR. RCAN[4] introduces an attention mechanism to further improve the reconstruction quality. SRFBN[2] introduces a feedback structure to optimize shallow features through iteration to produce deeper features.

The above MEF and SISR methods are used to solve the LDR and LR problems, respectively, but in real life, people often need to see HDR and HR images on cell phones or TVs, so the joint MEF and SR methods are necessary. This paper proposes a coupled feedback attention network-based image exposure fusion and super-resolution method, which can effectively suppress the superposition of redundant information in cyclic iterations, improve the quality of parameter sharing as well as exposure feature propagation.

2 Coupled Feedback Attention Network

In order to solve the propagation of redundant features and enhance the propagation of useful features in the coupled feedback network, this paper combines the coupled attention mechanism and feedback mechanism, and proposes an image exposure fusion and image super-resolution method based

AIoTC2022@International Conference on Artificial Intelligence, Internet of Things and Cloud Computing Technology
EMAIL: 2101440741@qq.com (Rong Wang), duanmu@zjnu.cn (Chunjiang Duanmu)
© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3351/paper02.pdf

on the coupled feedback attention network.

2.1 Basic network structure

The structure of the coupled feedback network is shown in Fig. 1. The shallow features F_{in}^{o} and F_{in}^{u} go through T rounds of iteration by the coupled feedback attention module in the upper and lower network, respectively. The feedback features in each iteration combine the feedback features in the other network and the shallow features in this network, together as the input of the next iteration, to achieve the refinement fused features. The coupled-feedback attention layer contains multiple coupled-feedback blocks and an attention module.

The extraction process of shallow features F_{in}^{o} and F_{in}^{u} of LR images can be expressed as

$$F_{in}^{o} = f_{FEB}(I_{lr}^{o})$$

$$F_{in}^{u} = f_{FEB}(I_{lr}^{u})$$

where f_{FEB} contains two convolutional layers Conv(3,4×m) and Conv(1,m), which are used to extract LR features and compress LR features, respectively. The extracted shallow features are first passed through SRB to obtain the deep features G^{o} and G^{o}, which can be expressed as

$$G^{o} = f_{SRB}(F_{in}^{o})$$

$$G^{o} = f_{SRB}(F_{in}^{u})$$

where f_{SRB} is the super-resolution module (SRB) operation.

Next, the deep exposure features of the two sub-networks are deeply fused after several iterations. At each iteration, the feedback features of the previous iteration are coupled and the shallow features F_{in}^{o} and F_{in}^{u} of the respective networks are together as the input of this iteration, and the feedback features C_{t}^{o} and C_{t}^{u} of the t-th iteration can be expressed as

$$C_{t}^{o} = f_{CFAB}(F_{in}^{o}, G_{t-1}^{o}, G_{t-1}^{u})$$

$$C_{t}^{u} = f_{CFAB}(F_{in}^{o}, G_{t-1}^{u}, G_{t-1}^{o})$$

where f_{CFAB} is the operation of the coupled feedback attention module. At the first iteration, G_{t-1}^{o} and G_{t-1}^{u} are the outputs G^{o} and G^{u} of the SRB, respectively.

Finally, the output of the coupled feedback attention module of each iteration and super-resolution features after channel attention module is reconstructed by the reconstruction module REC to obtain the SR residual image, then summed with the up-sampling of the corresponding LR image to produce the SR image:

$$I_{t}^{o} = f_{REC}(C_{t}^{o}) + f_{Up}(I_{lr}^{o})$$

$$I_{t}^{u} = f_{REC}(C_{t}^{u}) + f_{Up}(I_{lr}^{u})$$

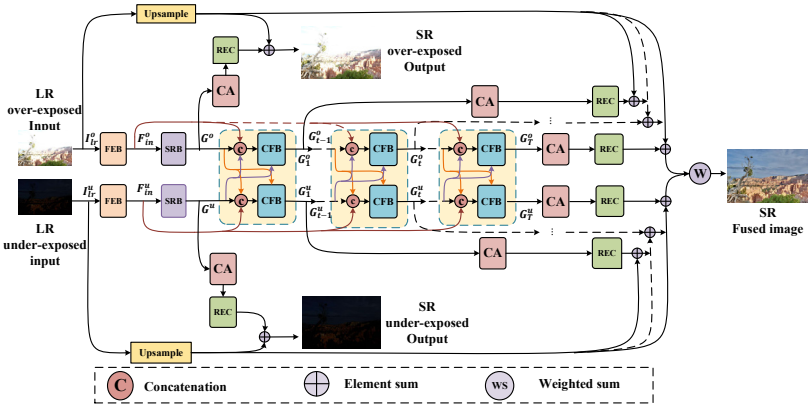

Figure 1 Coupled feedback attention network

2.2 Coupled Feedback Attention Module

This section specifically describe the specific iterative process of the coupled feedback block and channel attention module.

As shown in Fig. 2, the coupled feedback attention structure mainly contains iterative convolutional and deconvolutional layers constituting the CFB, and channel attention gates.

According to 3.1, in the upper sub-network, the inputs of the coupled feedback attention module are G_t^o, G_t^u, F_{in}^o. firstly, the channel compression is performed through the convolutional layer Conv(1,m) to obtain the input $L_t^o(0)$ of the coupled feedback attention module.

$$L_t^o(0) = f_{in}([G_t^o, G_t^u, F_{in}^o])$$

Next, multiple working groups consisting of convolutional and deconvolutional layers, the HR feature $H_t^o(n)$ of the n-th working group in the t-th iteration can be expressed as

$$H_t^o(n) = f_{Dec}([L_t^o(0), L_t^o(1), \dots, L_t^o(n-1)])$$

where $f_{Dec}$is the deconvolution layer Deconv(3,m). The HR features are generated by upsampling the LR features jointly from the first n-1 workgroups. Similarly, LR features $L_t^o(n)$ can be expressed as

$$L_t^o(n) = f_{Conv}([H_t^o(1), H_t^o(2), \dots, L_t^o(n-1)])$$

where f_{Conv} is the convolutional layer Conv(3,m).

The output of the final N-th working group is generated by the joint LR features of the previous N working groups passing through the convolution layer Conv(1,m) as follows.

$$G_t^o = f_{out}(L_t^o(1), L_t^o(2), \dots, L_t^o(N)])$$

The above describes the iterative process of the extreme high exposure branch, and the iterative process of the extreme low exposure branch is the same.

The feedback features G_t^o and G_t^u are output from each iteration, go through the channel attention module CA for feature optimization. The CA in this paper consists of three steps, which are global information compression, scaling and excitation, and recalibration.

1） Global information compression

In order to obtain the global information of each channel, this paper represents the feature values of each channel by global averaging pooling:

$$g_t^o = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} G_t^o(i,j)$$

$$g_t^u = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} G_t^u(i,j)$$

where $G_t^o(i,j)$ and $G_t^u(i,j)$ are the values at each position in the output extreme exposure feature, and compresses the multiple channels into a one-dimensional feature tensor.

2） Squeeze and excitation

In order to more fully explore the dependencies between individual channels, the paper introduces a gate mechanism for learning the nonlinear mapping between each channel and uses a sigmoid activation function to avoid the formation of adversarial relationships between channels, which can be expressed as

$$s_t^o = \sigma(W_2 \delta(W_1 g_t^o))$$

$$s_t^u = \sigma(W_2 \delta(W_1 g_t^u))$$

Where $W_1$and W_2 are the convolutional layer weights.

3） Recalibration

The original input features G_t^o individual channels are scaled by the channel attention weight matrix just learned, thus enhancing useful features and suppressing useless features:

$$C_t^o = \begin{cases} G_t^o \times (s_t^o + 1) & t = 1 \\ G_t^o \times s_{t-1}^o + G_t^o \times (s_t^o + 1) & t > 1 \end{cases}$$

$$C_t^u = \begin{cases} G_t^u \times (s_t^u + 1) & t = 1 \\ G_t^u \times s_{t-1}^u + G_t^u \times (s_t^u + 1) & t > 1 \end{cases}$$

Where s_t^o and s_t^u are the channel attention weights of the previous iteration.

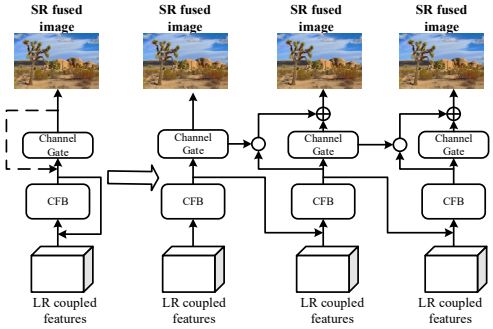

Figure 2 Coupled feedback attention structure

2.3 Loss Function

The method in this paper mainly achieves image super-resolution and image multi-exposure fusion, so the model in this paper uses a hierarchical loss function for optimization, and the loss function is expressed as

$$L_{total} = \lambda_o L_{SSIM}(I_{sr}^o, I_{gt}^o) + \lambda_u L_{SSIM}(I_{sr}^u, I_{gt}^u) + \sum_{t=1}^{T} \lambda_t (L_{SSIM}(I_t^o, I_{gt}) + L_{SSIM}(I_t^u, I_{gt}))$$

Where I_{gt}^o and I_{gt}^u are the HR standard images with extreme exposure, and I_{gt} is the HDR, HR standard image, which is the target to be achieved in the final fusion image. $\lambda_o, \lambda_u, \{\lambda_t\}_{t=1}^T$ are the weight coefficients of each loss part. In this paper, we set $\lambda_o = \lambda_u = \{\lambda_t\}_{t=1}^T = 1$.

3 Experiment and Analysis

3.1 Experiment Establishment

1）Experimental setup

In this paper, the training model was trained on GeForce GTX 1070Ti.The experiments in this paper mainly use SICE [5] dataset, which contains 589 high-quality reference images and their corresponding image sequences, and only extremely exposure are used in this paper.

2）Comparison Method

The network model proposed in this paper achieves both image super-resolution and image exposure fusion, we combine the current image super-resolution method and the image exposure fusion method as a comparison method. The image super-resolution methods are DBPN[3], RCAN[4], SRFBN[2], and SwinIR[9], and the main image exposure fusion methods are MGFF [10], FAST SPD-MEF [6], MEF-Net [8], and U2Fusion [7]. We combined SR methods and MEF methods, and changed the order of SR methods and MEF methods, i.e., SR+MEF or MEF+SR, to generate 32 comparison methods. The CF-Net [1] was also selected for comparison.

3.2 Objective evaluation

In order to verify the effectiveness of the method in this paper under magnification factor of 2, we use the SICE dataset and compare it with other advanced methods. These comparison methods are combined by SR method and MEF method. Table 1 shows the results of our method with the comparison methods for magnification factor of 2 under three metrics.

In Table 1, highlighting the first value of the fusion quality index in bold and the second ranked value in underline. From Table 1, we can see that the method of this paper has the best fusion effect, ranking first among 34 methods in metrics. PSNR index is improved by 0.25 dB, SSIM by 0.0028, and MEF-SSIM by 0.0005 compared to the second place CF-Net method.

Table 1. comparison of the fusion results under the magnification factor of 2

Super Resolution + Image Fusion												
Methods	MGFF[10]			FAST SPD-MEF[6]			MEF-Net[8]			U2Fusion[7]		
Combinations	PSNR	SSIM	MEF-SSIM	PSNR	SSIM	MEF-SSIM	PSNR	SSIM	MEF-SSIM	PSNR	SSIM	MEF-SSIM
DBPN[3]	17.47dB	0.7434	0.9121	17.30dB	0.7615	0.8976	17.26dB	0.7660	0.8888	17.83dB	0.7423	0.8807
RCAN[4]	17.39dB	0.7406	0.9114	17.34dB	0.7618	0.8974	17.24dB	0.7653	0.8882	17.85dB	0.7409	0.8804
SRFBN[2]	17.48dB	0.7425	0.9130	17.34dB	0.7601	0.8983	17.29dB	0.7641	0.8895	17.84dB	0.7402	0.8811
SWinIR[9]	17.44dB	0.7436	0.9113	17.26dB	0.7618	0.8968	17.23dB	0.7667	0.8881	17.82dB	0.7436	0.8802
Image Fusion + Super Resolution												
Methods	DBPN[3]			RCAN[4]			SRFBN[2]			SWinR[9]		
Combinations	PSNR	SSIM	MEF-SSIM	PSNR	SSIM	MEF-SSIM	PSNR	SSIM	MEF-SSIM	PSNR	SSIM	MEF-SSIM
MGFF[10]	17.27dB	0.7161	0.9144	17.18dB	0.7122	0.9135	17.38dB	0.7218	0.9158	17.19dB	0.7135	0.9131
Fast SPD-MEF[6]	17.26dB	0.7554	0.8954	17.24dB	0.7533	0.8949	17.31dB	0.7557	0.8962	17.21dB	0.7546	0.8944
MEF-Net[8]	17.25dB	0.7636	0.8886	17.23dB	0.7624	0.8882	17.27dB	0.7630	0.8892	17.20dB	0.7629	0.8878
U2Fusion[7]	17.81dB	0.7384	0.8843	17.82dB	0.7368	0.8837	17.85dB	0.7395	0.8850	17.76dB	0.7374	0.8835
CF-Net[1]	PSNR=<u>21.24</u>dB			SSIM=<u>0.8140</u>						MEF-SSIM=<u>0.9332</u>		
Ours	PSNR=**21.49**dB			SSIM=**0.8168**						MEF-SSIM=**0.9337**		

3.3 Subjective evaluation

Fig. 3 visually depicts the fused images produced by this paper and other advanced methods at magnification of factor 2. From the experimental results, it can be seen that compared with SR+MEF and MEF+SR methods, the method in this paper has a great improvement in details, and compared with the coupled feedback network, this paper alleviates the phenomenon that there is redundant information in the image due to the coupled feedback mechanism.

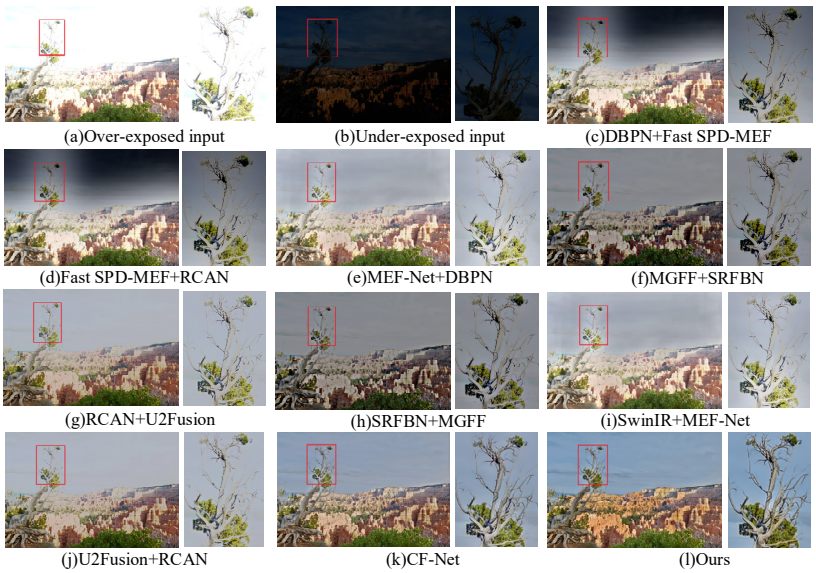

Figure 3 Comparison of different methods of "landscape" under 2×

4 Conclusion

Based on the powerful image reconstruction property of feedback mechanism and the property that channel attention mechanism can distinguish the importance of features. In this paper, a coupled feedback attention network is proposed to solve the image super-resolution problem and image exposure fusion problem simultaneously. The experimental results show that the algorithm in this paper retains

the detailed information of edges, region boundaries and textures of the original image sequence.

5 References

[1] Deng X., Zhang Y. T., Xu M., et al. Deep coupled feedback network for joint exposure fusion and image super-resolution[J]. IEEE Transactions on Image Processing. 2021, 30:3098-3112.

[2] Li Z., Yang J. L., Liu Z., et al. Feedback Network for Image Super-Resolution[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

[3] Harris M., Shakhnarovich G., Ukita N., et al. Deep back-projectinetworks for single image super-resolution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2021, 43(12):4323-4337.

[4] Zhang T. L., Li K. P., Li K., et al. Image Super-Resolution Using Very Deep Residual Channel Attention Networks[C]. European Conference on Computer Vision. 2018.

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

[6] Li H., Ma KD., Yong H. W., et al. Fast multi-scale structural patch decomposition for multi-exposure image fusion[J]. IEEE Transactions on Image Processing. 2020, 29: 5805-5816.

[7] Xu X., Ma J. Y., Jiang J. J., et al. U2Fusion: A unified unsupervised image fusion network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022, 44(1): 502-518.

[8] Ma K., Duanmu Z., Zhu H., Fang Y., et al. Deep guided learning for fast multi-exposure image fusion[J]. IEEE Transactions on Image Processing. 2020, 29: 2808–2819.

[9] Liang J. Y., Cao J. Z., Sun G. L., et al. SwinIR: Image Restoration Using Swin Transformer[C]. IEEE/CVF International Conference on Computer Vision Workshops. 2021. Electr Network.

[10] Bavirisetti D. P., Xiao G., Zhao J. H., et al. Multi-scale guided image and video fusion: a fast and efficient approach[J]. Circuits Systems and Signal Processing. 2019, 38(12): 5576-5605.