

# Efficient Multi-exposure Image Fusion via Filter-dominated Fusion and Gradient-driven Unsupervised Learning, Supplementary Material

Kaiwen Zheng Jie Huang Hu Yu Feng Zhao\*

University of Science and Technology of China

{kezh, hj0117, yuhu520}@mail.ustc.edu.cn, fzhao956@ustc.edu.cn

## 1. Hierarchical Representation Extractor (HRE)

In this section, we supplement the detailed architecture of the proposed HRE for hierarchical representation extraction. As shown in Fig. 1, HRE is designed in a Unet-like fashion with Resblock and convolutional layer as basic components. Then, the feature map calculated from different levels is able to act as the hierarchical representation of the source image. Besides, several flexible skip connections are employed to better incorporate multi-level features for elevating the network’s mapping capabilities. Specifically, given the feature maps ( $F_{base}$ ,  $F_{skip}$ ) from the current layer and the skip connection, the fusion procedure is as follows:

$$F_{fused} = \alpha \times Conv_{1*1}([F_{base}, F_{skip}]) + \beta \times F_{base}. \quad (1)$$

where  $\alpha$  and  $\beta$  are learnable parameters, and  $[\cdot]$  means concatenation operation.

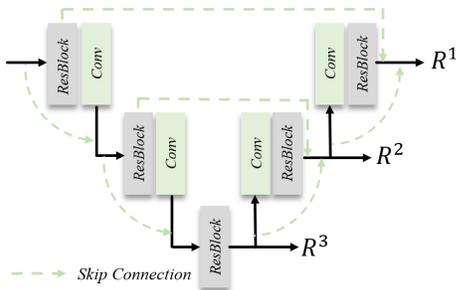


Figure 1. The overall architecture of the Hierarchical Representation Extractor (HRE).

## 2. Supplemental Ablation Studies

**Compare to previous fusion paradigms.** To demonstrate the effectiveness of our proposed fusion paradigm, we conducted comparative experiments with the existing fusion paradigms. We ensured that the models had comparable parameter scales and identical settings. Fig. 2 illustrates the

\*Corresponding author.

visual comparison results, which confirm that our paradigm achieves superior fusion performance.

Furthermore, we performed an ablation study on the effect of the dilation operation in GIF loss and reported the results in Fig. 2. The comparison reveals that omitting the dilation operation leads to a slightly lower brightness in the fusion result, but it does not substantially degrade the fusion quality.

## 3. Visualization

In this section, more qualitative results are supplemented for a comprehensive comparison. In Fig. 3, we supplement the visual comparison of multi-exposure fusion by comparing six traditional methods and nine deep learning-based methods with our method. Most traditional approaches cannot reach a globally consistent exposure level, resulting in inferior visual effects, but have higher information fidelity (see outdoor scenes). In contrast, methods based on deep learning are the exact reverse. Even though our method has produced excellent perceptual results, there is still room for development to achieve satisfactory multi-exposure fusion results.

In addition, Fig. 4 provides a supplement to the extended experimental results on Multi-Focus image Fusion (MFF) and Visible-Infrared image Fusion (VIF). It can be seen that in the MFF task, our method is able to generate global sharpening results by fusing near-focus and far-focus images. In the VIF process, infrared imaging can provide an essential complement to visible images in extreme environments. Extensive visualization results demonstrate the feasibility of our method as a general image fusion paradigm.

## 4. Metrics

In this section, we supplement the definitions of the evaluation metrics used in the quantitative experiments. All indicators are calculated in the same way as used in the benchmark [14]. In the following definitions,  $H$  and  $W$  represent the height and width of the image. The source image



Figure 2. Visualization of the ablation results over previous fusion paradigms and the dilation process within the GIF. The weight map-based fusion paradigm and the deep-represented fusion paradigm are marked as baseline1 and baseline2, respectively.

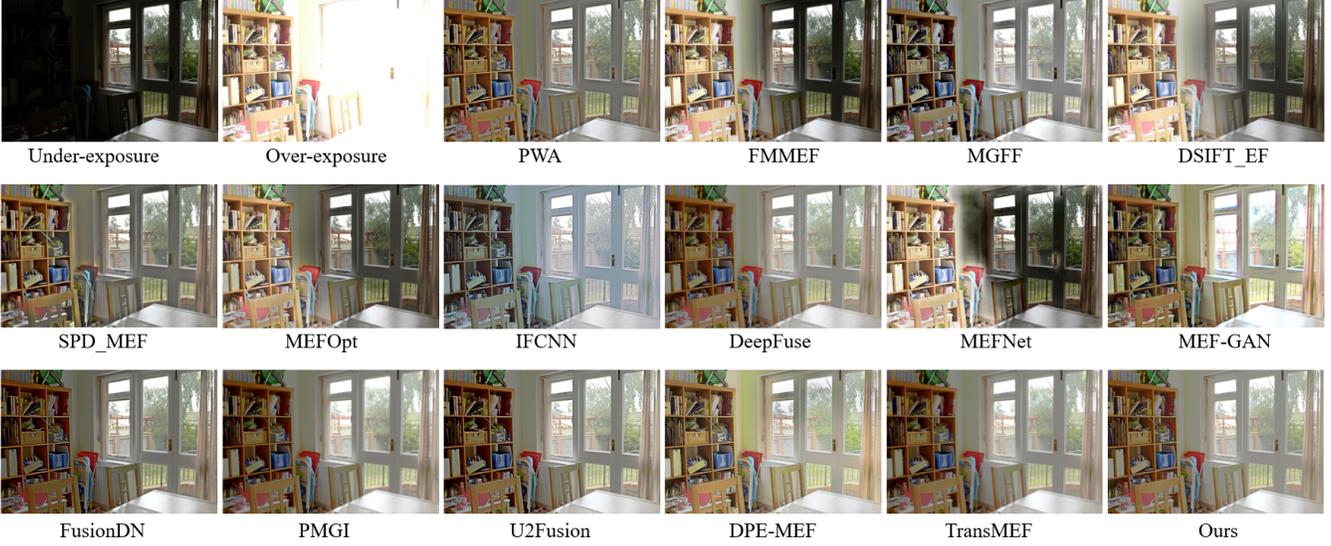


Figure 3. Supplementary presentation of qualitative experiments on the multi-exposure fusion task.

pair and fused image are denoted by  $(I_0, I_1)$  and  $I_f$ , respectively.

**Cross entropy (CE).** CE [1] is computed as:

$$CE = \frac{(\sum_{i=0}^{255} h_{I_0}(i) \log_2 \frac{h_{I_0}(i)}{h_{I_f}(i)} + \sum_{i=0}^{255} h_{I_1}(i) \log_2 \frac{h_{I_1}(i)}{h_{I_f}(i)})}{2}, \quad (2)$$

where  $h(\cdot)$  is the normalized histogram of the image.

**Entropy (EN).** EN [11] evaluates the information richness of the fused image and is defined as follows:

$$EN = - \sum_{l=0}^{L-1} p_l \log_2 p_l, \quad (3)$$

where  $L$  is the number of gray levels and  $p_l$  is the normalized histogram of the corresponding gray level in the fused image.

**Feature mutual information (FMI).** FMI [5] represents the correlation between source image and fused image in feature information and is defined as follows:

$$FMI = MI_{\hat{I}_0, \hat{I}_f} + MI_{\hat{I}_1, \hat{I}_f}, \quad (4)$$

where  $\hat{I}_0, \hat{I}_1, \hat{I}_f$  are the feature maps of  $I_0, I_1, I_f$ . MI denotes the mutual information calculation [9].

**Normalized mutual information (NMI).** NMI [7] is defined as:

$$NMI = 2 \left( \frac{MI_{I_0, I_f}}{H(I_0) + H(I_f)} + \frac{MI_{I_1, I_f}}{H(I_1) + H(I_f)} \right), \quad (5)$$

where  $H(\cdot)$  stands the image entropy.

**Peak signal-to-noise ratio (PSNR).** PSNR [8] is defined as:

$$PSNR = 10 \log_{10} \frac{m^2}{MSE}, \quad (6)$$

where  $m$  is the max value of the fused image,  $MSE$  is the mean squared error between the fused image and source images.

**Nonlinear correlation information entropy ( $Q_{NCIE}$ ).** First, a nonlinear correlation matrix  $R$  based on nonlinear correlation coefficient ( $NCC$ ) between source images and the fused image is obtained as:

$$R = \begin{bmatrix} 1 & NCC_{I_0, I_1} & NCC_{I_0, I_f} \\ NCC_{I_1, I_0} & 1 & NCC_{I_1, I_f} \\ NCC_{I_f, I_0} & NCC_{I_f, I_1} & 1 \end{bmatrix}. \quad (7)$$

Then,  $Q_{NCIE}$  [12], can be computed as:

$$Q_{NCIE} = 1 + \sum_3^{i=1} \frac{\lambda_i}{3} \log_{256} \frac{\lambda_i}{3}, \quad (8)$$

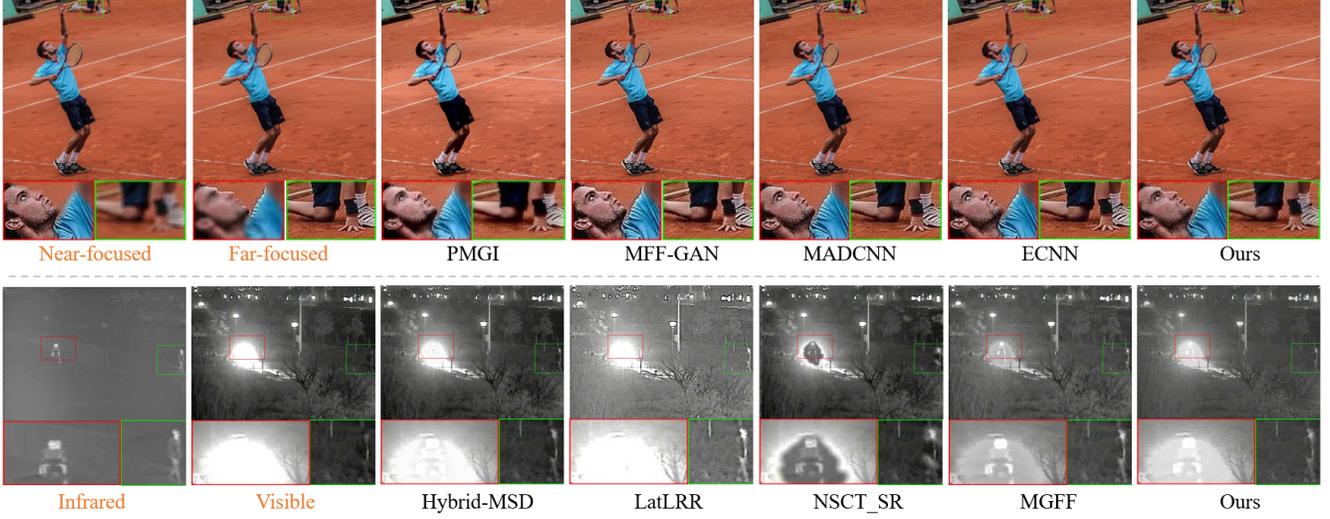


Figure 4. Supplementary presentation of qualitative experiments on the multi-focus fusion and visible-infrared fusion task.

where  $\lambda_i$  are the eigenvalues of the matrix  $R$ .

**Average gradient (AG).** AG [3] measures the texture richness of the fused image through gradient information as follows:

$$AG = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \sqrt{\frac{\nabla F_x^2(i,j) + \nabla F_y^2(i,j)}{2}}, \quad (9)$$

where  $\nabla F_x(i,j) = I_f(i,j) - I_f(i+1,j)$ ,  $\nabla F_y(i,j) = I_f(i,j) - I_f(i,j+1)$ .

**Edge intensity (EI).** EI [10] measures the edge information of the fused image through the Sobel gradient operator as:

$$EI = \sqrt{(I_f * h_x)^2 + (I_f * h_y)^2}, \quad (10)$$

where  $h_x, h_y$  represent the horizontal and vertical operators, respectively.

**Edge based similarity measurement ( $Q^{AB/F}$ ).** ( $Q^{AB/F}$ ) [13] is defined as:

$$Q^{AB/F} = \frac{\sum_{i=1}^H \sum_{j=1}^W (Q^{I_0, I_f} w^{I_0} + Q^{I_1, I_f} w^{I_1})}{\sum_{i=1}^H \sum_{j=1}^W (w^{I_0} + w^{I_1})}, \quad (11)$$

where  $Q^{I_i, I_f} = Q_g^{I_i, I_f} Q_a^{I_i, I_f}$ ,  $Q_g^{I_i, I_f}$  and  $Q_a^{I_i, I_f}$  represents the edge strength and orientation values, respectively.  $w^X$  denotes the weight that expresses the importance of each source image to the fused image in gradient domain.

**Spatial frequency (SF).** SF [4] is defined as:

$$SF = \sqrt{RF^2 + CF^2}, \quad (12)$$

where  $RF = \sqrt{\sum_{i=1}^H \sum_{j=1}^W (I_f(i,j) - I_f(i,j-1))^2}$  and  $CF = \sqrt{\sum_{i=1}^H \sum_{j=1}^W (I_f(i,j) - I_f(i-1,j))^2}$  is generated to measure the gradient distribution of fused image.

**Human visual perception ( $Q_{CB}$ ).**  $Q_{CB}$  [2] is defined as:

$$Q_{CB} = \frac{\lambda_{I_0} Q_{I_0, I_f} + \lambda_{I_1} Q_{I_1, I_f}}{HW}, \quad (13)$$

where  $Q_{I_0, I_f}$  and  $Q_{I_1, I_f}$  indicates the contrast transformed from source images to the fused image.  $\lambda_{I_0}$  and  $\lambda_{I_1}$  are the saliency maps of  $Q_{I_0, I_f}$  and  $Q_{I_1, I_f}$ , respectively.

**Visual information fidelity (VIF).** VIF measures the distortion between the fused image and the source images to measure visual information fidelity. The computational process can be found in previous study [6].

## 5. Broader Impact

Incomplete information capture is the most common and unavoidable phenomenon in single-shot imaging of commonly used imaging sensors, such as underexposure, overexposure, and partial focus. Therefore, image fusion technology has broad impacts and practical values in various applications, including remote sensing, medicine, astronomy, military, and civilian imaging equipment. Image fusion technology aims to generate high-quality and information-rich images from given under-expected counterparts. For a wider range of application scenarios with limited computing resources, the lower model complexity and running time deserve more valuable evaluation. To this end, our efficient and high-fidelity fusion method based on the proposed filter-dominated fusion paradigm can provide high-quality fused images to facilitate intelligent data analysis tasks in these fields.

The negative consequences may accompany image fusion technology despite the many benefits it brings. This is mainly related to the risks that may arise from the authenticity of the fused images. For instance, the fusion of images of the same site captured at different times may result in the

creation of convincing fake photos, resulting in detrimental ramifications. Furthermore, it is essential to be aware of the possible failure results of any image fusion algorithm, leading to misjudgment of information.

## References

- [1] DM Bulanon, TF Burks, and V Alchanatis. Image fusion of visible and thermal images for fruit detection. *Biosystems Engineering*, 103(1):12–22, 2009. 2
- [2] Yin Chen and Rick S Blum. A new automated quality assessment algorithm for image fusion. *Image and Vision Computing*, 27(10):1421–1432, 2009. 3
- [3] Guangmang Cui, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition. *Optics Communications*, 341:199–209, 2015. 3
- [4] Ahmet M Eskicioglu and Paul S Fisher. Image quality measures and their performance. *IEEE Transactions on Communications*, 43(12):2959–2965, 1995. 3
- [5] Mohammad Bagher Akbari Haghghat, Ali Aghagolzadeh, and Hadi Seyedarabi. A non-reference image fusion metric based on mutual information of image features. *Computers & Electrical Engineering*, 37(5):744–756, 2011. 2
- [6] Yu Han, Yunze Cai, Yin Cao, and Xiaoming Xu. A new image fusion performance metric based on visual information fidelity. *Information Fusion*, 14(2):127–135, 2013. 3
- [7] M Hossny, S Nahavandi, and D Creighton. Comments on "information measure for performance of image fusion". *Electronics Letters*, 44(18):1066–1067, 2008. 2
- [8] P Jagalingam and Arkal Vittal Hegde. A review of quality metrics for fused image. *Aquatic Procedia*, 4:133–142, 2015. 2
- [9] Guihong Qu, Dali Zhang, and Pingfan Yan. Information measure for performance of image fusion. *Electronics Letters*, 38(7):1, 2002. 2
- [10] B Rajalingam and R Priya. Hybrid multimodality medical image fusion technique for feature enhancement in medical diagnosis. *International Journal of Engineering Science Invention*, 2(Special issue):52–60, 2018. 3
- [11] J Wesley Roberts, Jan A Van Aardt, and Fethi Babikker Ahmed. Assessment of image fusion procedures using entropy, image quality, and multispectral classification. *Journal of Applied Remote Sensing*, 2(1):023522, 2008. 2
- [12] Qiang Wang, Yi Shen, and Jian Qiu Zhang. A nonlinear correlation measure for multivariable data set. *Physica D: Nonlinear Phenomena*, 200(3-4):287–295, 2005. 2
- [13] Costas S Xydeas, Vladimir Petrovic, et al. Objective image fusion performance measure. *Electronics Letters*, 36(4):308–309, 2000. 3
- [14] Xingchen Zhang. Benchmarking and comparing multi-exposure image fusion algorithms. *Information Fusion*, 74:111–131, 2021. 1