



Article

MEFSR-GAN: A Multi-Exposure Feedback and Super-Resolution Multitask Network via Generative Adversarial Networks

Sibo Yu ^{1,2} , Kun Wu ¹, Guang Zhang ¹, Wanhong Yan ¹, Xiaodong Wang ¹ and Chen Tao ^{1,*}

¹ Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China; yusibo@ciomp.ac.cn (S.Y.); wukun@ciomp.ac.cn (K.W.); zhangguang0920@ciomp.ac.cn (G.Z.); yanwanhong@ciomp.ac.cn (W.Y.); wangxiaodong@ciomp.ac.cn (X.W.)
² University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: taochen@ciomp.ac.cn

Abstract: In applications such as satellite remote sensing and aerial photography, imaging equipment must capture brightness information of different ground scenes within a restricted dynamic range. Due to camera sensor limitations, captured images can represent only a portion of such information, which results in lower resolution and lower dynamic range compared with real scenes. Image super resolution (SR) and multiple-exposure image fusion (MEF) are commonly employed technologies to address these issues. Nonetheless, these two problems are often researched in separate directions. In this paper, we propose MEFSR-GAN: an end-to-end framework based on generative adversarial networks that simultaneously combines super-resolution and multiple-exposure fusion. MEFSR-GAN includes a generator and two discriminators. The generator network consists of two parallel sub-networks for under-exposure and over-exposure, each containing a feature extraction block (FEB), a super-resolution block (SRB), and several multiple-exposure feedback blocks (MEFBs). It processes low-resolution under- and over-exposed images to produce high-resolution high dynamic range (HDR) images. These images are evaluated by two discriminator networks, driving the generator to generate realistic high-resolution HDR outputs through multi-goal training. Extensive qualitative and quantitative experiments were conducted on the SICE dataset, yielding a PSNR of 24.821 and an SSIM of 0.896 for $2\times$ upscaling. These results demonstrate that MEFSR-GAN outperforms existing methods in terms of both visual effects and objective evaluation metrics, thereby establishing itself as a state-of-the-art technology.

Keywords: multitask networks; super resolution; multiple exposure fusion; generative adversarial networks



Citation: Yu, S.; Wu, K.; Zhang, G.; Yan, W.; Wang, X.; Tao, C. MEFSR-GAN: A Multi-Exposure Feedback and Super-Resolution Multitask Network via Generative Adversarial Networks. *Remote Sens.* **2024**, *16*, 3501. <https://doi.org/10.3390/rs16183501>

Academic Editors:

Fahimeh Farahnakian,
Jukka Heikkonen, Pouya Jafarzadeh
and Farshad Farahnakian

Received: 24 July 2024

Revised: 7 September 2024

Accepted: 13 September 2024

Published: 21 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In natural scenes, objects can have significant differences in brightness and thus, a wide dynamic range. However, camera sensors have limitations that prevent them from capturing the full range of information [1]. As a result, captured images often exhibit distortion, noise, and other issues, thus leading to lower resolution compared with real scenes [2]. Common methods employed to address this issue include single image super resolution (SISR) and multiple-exposure fusion (MEF).

Single image super resolution is a significant challenge within the field of computer vision. It focuses on enhancing the quality of low-resolution (LR) images to produce clear and detailed high-resolution (HR) versions [3]. This technology holds significant value in aerospace and remote sensing image processing, where captured images often appear to be low-resolution due to equipment constraints. SISR techniques play a vital role in overcoming these limitations and enable the generation of high-quality images that offer enhanced data support for diverse applications in remote sensing analysis [4]. This is crucial for enhancing information extraction capabilities and analytical depth in the aforementioned fields. Previous approaches [5,6] initially treated SISR as an interpolation

issue, resulting in fast processing but often sacrificing high-frequency details. More recent efforts [7–10] have focused on learning degradation models from unpaired real image data to enhance generalization. However, these learning-based techniques rely heavily on training data and may exhibit significant performance decline when encountering unforeseen degradations during testing. Leveraging the powerful representation capabilities of convolutional neural networks (CNNs), several methods [11,12] have emerged that utilize deep learning to map low-resolution (LR) to high-resolution (HR) images. In recent years, methods based on implicit neural representations, such as LIIF and FunSR, have led to significant breakthroughs in super-resolution tasks. Unlike traditional explicit representation methods, implicit neural representation techniques model image content by learning continuous functions, enabling the generation of high-resolution images at any scale while retaining greater detail and offering enhanced flexibility. These methods are particularly effective in the realm of remote sensing image processing, where they address challenges associated with multi-scale variations and complex scenes, thereby improving both the reconstruction quality and computational efficiency of images. These have achieved state-of-the-art results. For instance, to tackle potential blur and distortion issues in the produced super-resolution images, SRGAN [13] introduced a method based on generative adversarial networks (GANs). Through an adversarial training mechanism, the Super Resolution Generative Adversarial Network (SRGAN) is capable of generating realistic high-resolution images and effectively restoring intricate details. Particularly in image super-resolution tasks, the GAN framework excels at capturing and reconstructing subtle texture features through the adversarial learning process that occurs between its generator and discriminator, thereby surpassing the performance of traditional methods. The network discussed in the present paper also adopts the GAN framework.

Multiple-exposure image fusion involves merging multiple low-dynamic-range (LDR) images with varying exposure levels to create high-dynamic-range (HDR) images [14]. Due to hardware or optical limitations, a single sensor can capture only a portion of the available information, such as the brightness of reflected light within a specific range and depth of field. Image fusion aims to combine data from multiple images taken using different sensors or setups to create a composite image with enhanced scene representation and improved visual perceptibility. This process is particularly valuable in aerospace and remote sensing applications, where capturing a wide range of brightness information within a limited dynamic range is crucial for tasks like satellite remote sensing and aerial photography. MEF technology effectively addresses the limitations of LDR images by combining multiple images taken under varying exposure conditions to create HDR images [15]. In particular, GAN-based multi-exposure fusion technology effectively integrates image information captured under varying exposure conditions to produce high-dynamic-range (HDR) images, preserving intricate details and realistic lighting effects. These attributes position GAN as a potent tool for multi-exposure fusion tasks. This process significantly improves the image's dynamic range and detail performance. For instance, in satellite remote sensing, MEF can merge LDR images captured at different times to enhance ground object information [16]. MEF excels in capturing details in dark or reflective areas and offers high-quality data for geographic information extraction and environmental monitoring. Similarly, in aerial photogrammetry, MEF is valuable for producing clear and detailed HDR orthophotos by blending images with different exposures. These enriched data are beneficial for applications like 3D modeling, change detection, urban planning, and infrastructure management. Furthermore, in astronomical imaging, MEF can fuse starscape images taken at varying exposure times to preserve dark details and capture detailed textural information of bright objects. This approach provides valuable data for astronomical research. In general, MEF technology holds significant potential in aerospace remote sensing since it expands the dynamic range of imaging equipment to produce detailed HDR images for various aerospace applications [17]. The continuous advancement of MEF methods will further enhance their value in the aerospace sector. These methods can be categorized as either non-extreme exposure fusion or extreme exposure fusion, based on the number of

LDR images used. In non-extreme cases, fusion performance is closely tied to the quantity of LDR images available [15]. While a greater number of LDR images typically results in better fusion outcomes, it also increases storage and computational complexity, and it may even be impossible to obtain a large number of LDR images. To address these limitations, recent advancements in fusion methods have introduced extreme fusion techniques that rely on just a pair of extreme exposure images to enhance fusion outcomes [18]. Given that extreme exposure images typically contain implicit information, these methods leverage deep convolutional neural networks to thoroughly analyze and merge this information. The present study also belongs to this realm of extreme fusion techniques.

As confirmed in previous research [19], HDR images have been shown to offer more enhanced features in comparison to LDR images. Additionally, recent findings [20] indicate that super-resolution techniques can greatly enhance the precision of object detection. To achieve high-quality performance in diverse tasks related to remote sensing imaging and astronomical observation, it is essential to simultaneously apply SR and MEF processing to remote sensing or astronomical images to generate HR images with HDR capabilities. Despite the abundance of studies on SR and MEF, these topics are typically treated as separate research inquiries. Xin Deng [21] introduced a coupled feedback network (CF-Net): a deep neural network combining MEF and SR tasks in an end-to-end CNN framework. In CF-Net, the synergy between MEF and SR tasks is prioritized since better fusion results can enhance SR accuracy, which in turn can boost fusion performance. Inspired by CF-Net, our present work incorporates a GAN architecture to process two low-resolution over-exposed images and generate HDR-SR images simultaneously, thus leading to significant enhancements in MEF and SR performance. Notably, our work presents the first end-to-end GAN framework for achieving state-of-the-art results in both MEF and SR tasks simultaneously. The key contributions of this research are outlined below:

1. Introduction of an end-to-end Multi-Exposure Super-Resolution Generative Adversarial Network (MEFSR-GAN): This paper presents the first use of a GAN framework to simultaneously achieve multi-exposure fusion (MEF) and super resolution (SR) in a unified model. MEFSR-GAN effectively enhances both MEF and SR performance under extreme exposure conditions;
2. Development of a multi-exposure feedback block (MEFB): We propose a novel MEFB specifically designed to handle low-resolution images with over-exposure and under-exposure. The MEFB processes highly exposed images in parallel and incorporates a channel attention mechanism to optimize feature extraction and improve model generalization;
3. Proposal of a dual discriminator network: To tackle the challenges of training on extremely exposed images, we introduce a dual discriminator network that guides the generator to learn stable and distinct feature representations, producing images that closely resemble the ground truth.
4. State-of-the-art results with SICE and PQA-MEF datasets: Experimental results obtained using the SICE and PQA-MEF datasets demonstrate that MEFSR-GAN outperforms the latest MEF and SR methods, achieving state-of-the-art performance in both qualitative and quantitative evaluations.

2. Related Work

In this section, we review related work on super-resolution image reconstruction and multi-exposure image fusion.

2.1. Single-Image Super-Resolution

SISR is a technique that focuses on generating HR images from LR images to produce detailed and natural results [22]. This field has gained significant attention recently due to its practical applications. SISR methods involve mapping LR images to HR counterparts, and degradation models determine how HR images are transformed into LR images [23]. Two common degradation models are (1) bicubic degradation, which uses bicubic interpo-

lation to generate LR images, and (2) traditional degradation, which can be mathematically modeled as follows:

$$y = (x \otimes k) \downarrow_s + n, \quad (1)$$

The process of obtaining an LR image involves convolving an HR image with a Gaussian kernel (or point spread function) k to create a blurred image y . This is followed by downsampling operation \downarrow_s with a scale factor s and the addition of white Gaussian noise with standard deviation σ . Bicubic degradation is seen as a special case of traditional degradation since it can be approximated by adjusting the kernel with zero noise. Degradation models are typically defined by factors like blur kernels and noise level. Depending on prior knowledge of these factors, SISR methods are categorized into non-blind methods and blind methods.

Single-image super resolution is considered an ill-posed inverse problem because one LR image can be associated with multiple HR images [24]. In 1991, Irani and Peleg proposed the iterative back projection (IBP) method [25]. In 1992, Ur and Gross put forward the non-uniform difference method to enhance image resolution [26]. In the same year, Schulz and Stevenson introduced the maximum a posterior probability (MAP) method [27,28]. Elad and Feuer proposed the adaptive filtering method in 1999 [29]. Then, in 2002, Lertrattanapanich presented the Delaunay triangulation network reconstruction method [30] along with enhanced algorithms and joint algorithms that incorporated various regularization terms.

Recent advances in artificial intelligence have propelled learning-based super-resolution reconstruction to the forefront, with foundational work by Freeman in 2000 [31]. Subsequently, several influential approaches have emerged, including the sparse representation method introduced by Yang in 2008 [32,33]; the neighbor-model-based domain restoration (anchored neighborhood regression (ANR)) method proposed by Timofte in 2013 [34]; and the deep convolutional neural network-based method (SRCNN) introduced by Dong at the Chinese University of Hong Kong in 2014 [35], which significantly advanced the field of super resolution.

In the field of SISR reconstruction, the advancement of deep learning techniques has led to the creation of various novel network architectures. Each of these architectures is designed with unique features and strategies to enhance the quality of reconstructed images. This section provides an overview and assessment of several notable networks such as SRCNN [35], VDSR [36], DRCN [37], SRGAN [13], MemNet [38], EDSR [39], RCAN [40], HAT [41], SwinIR [42], and SAN [43].

SRCNN [35] performs image reconstruction tasks using a three-layer convolutional neural network. Its simplicity and efficiency have paved the way for further research. However, due to its shallow nature, SRCNN has limitations in processing complex textures. This indicates the need for deeper network structures to capture more intricate image features.

VDSR [36] builds on SRCNN by deepening the network to 20 convolutional layers and introducing residual learning, which accelerates convergence and improves super-resolution performance. However, the increased training complexity and computational demands of VDSR emphasize the need to balance efficiency and performance.

DRCN [37] introduces a recursive convolutional structure, reusing layers multiple times to reduce the number of parameters. It also employs multi-supervised training with supervision signals at various depths to enhance efficiency. However, the recursive nature adds computational complexity.

SRGAN [13] integrates generative adversarial networks with residual networks, using stacked residual blocks to extract high-level semantic features. A discriminator guides the generator to produce more realistic HR images, though potential for artifacts exists.

EDSR [39] enhances SRGAN by removing the BatchNorm layer, adding residual blocks, and increasing feature channels. It also introduces multi-scale fusion for handling varying scale information. Despite these improvements, EDSR has extensive model parameters, leading to high computational demands.

RCAN [40] builds on EDSR by incorporating a channel attention mechanism to dynamically adjust feature weights, prioritizing those beneficial for super-resolution tasks. It uses a residual grouping strategy to increase network depth, achieving top-tier performance but with notable model complexity.

MemNet [38] introduces memory blocks to extract features at different receptive field scales, improving the ability to process multi-scale features. However, its high complexity presents training challenges.

Meanwhile, SwinIR [42] employs depthwise separable convolution to reduce computational load and incorporates adversarial and perceptual losses to enhance visual quality. The transformer architecture excels at capturing long-range dependencies and multi-scale features, producing images with superior detail. However, SwinIR's large model parameters and computational requirements, along with potentially weaker performance when used with very small LR images, are notable drawbacks.

In recent years, the aforementioned methods have employed deep network structures to capture multi-scale features of images, thereby preserving more details during the image reconstruction process. However, these techniques typically depend on fixed magnification factors and lack the flexibility needed for super resolution of images at arbitrary scales. The local implicit image function (LIIF) [43] is a technology developed to address the challenge of image super resolution at any scale. LIIF predicts the corresponding pixel value based on any input coordinates by learning a continuous mapping function that relates image coordinates to pixel values. This method does not rely on fixed interpolation operations, allowing it to generate super-resolution images at any scale and making it particularly well-suited for processing remote sensing image data characterized by varying resolutions and intricate details.

The various networks discussed above exhibit distinct strengths and limitations, hence showcasing a wide range of exploration and advancement in the realm of super-resolution reconstruction. Subsequent research should aim to strike a more optimal balance between enhancing image reconstruction quality, minimizing computational expenses, and enhancing model robustness.

2.2. Multi-Exposure Fusion

Multi-exposure fusion is a significant area of research focused on producing HDR images by combining multiple images of the same scene captured at varying exposure levels. Conventional techniques relying on the Laplacian pyramid have evolved into approaches utilizing deep learning and offering unique features and applications.

Since the initial proposal by Burt and Adelson in 1983 to utilize the Laplacian pyramid for image fusion, it has been fundamental in MEF research [44]. Mertens et al. [45] introduced the first pixel-level MEF method within the Laplacian pyramid framework, and they successfully balanced visual quality and computational complexity. Subsequent pixel-level MEF methods built upon this framework aimed to enhance visual quality, albeit often at the cost of increased computational demands [46]. Burt et al. [47] developed weights based on local energy and the correlation between pyramids, effectively reducing artifact generation; this marked a significant advancement in MEF research.

Compared with pixel-level methods, patch-based approaches generate smoother weight maps but at a higher computational cost. Goshtasby [48] and Ma and Wang [49] demonstrated the potential of non-overlapping block-based fusion, though block artifacts remain an issue. Techniques like pixel-based and multi-scale transformations (e.g., pyramids, wavelet transforms) improve visual quality but struggle with accurately representing curves and edges, highlighting the complexities in multi-spectral and multi-modal fusion.

To reduce artifacts in dynamic scenes, Liu and Wang [50] introduced dense SIFT (DSIFT) for improved image alignment, enhancing quality but at the cost of high computational demands and complex parameter tuning. Sang-hoon Lee et al. [51] developed an adaptive weight-based method combining pixel intensity and gradient information

to reduce halo and ghost effects, though challenges persisted under extreme exposure conditions and real-time processing.

Recently, deep learning-based multi-exposure fusion has gained traction. DeepFuse [52] pioneered this integration with a focus on structure and contrast, though it may overlook other important information. IFCNN [53] and MEF-CNN [54] offer CNN-based frameworks with strong visual effects but require significant optimization and high computational resources. MEF-GAN [55] employs GANs to enhance naturalness, while FusionDN [56] uses densely connected networks for detailed fusion, though both are computationally intensive.

Xu's U2Fusion [57] introduced a unified unsupervised network to merge image information without manual annotations, enhancing adaptability and automation. This approach is especially suited for scenarios like night vision and infrared fusion, but achieving accuracy comparable to supervised methods can be challenging due to the reliance on large amounts of unlabeled data.

As shown through the above review, SR and MEF tasks are commonly considered separate research inquiries. This paper makes a significant contribution to the literature by introducing an end-to-end network utilizing a GAN framework to accomplish both image fusion and super resolution simultaneously.

3. Multi-Exposure Feedback and Super-Resolution Generative Adversarial Networks (MEFSR-GAN)

In this section, we provide a detailed introduction to the proposed MEFSR-GAN. We begin by outlining the overall network architecture in Section 3.1. Next, in Section 3.2, we delve into the unique architecture of the MEFB utilized in our network. Then, in Section 3.3, we discuss the discriminator network. Finally, in Section 3.4, we discuss the employed loss function.

3.1. Network Architecture

Our proposed MEFSR-GAN features a GAN architecture with a generator and two discriminators, as shown in Figure 1. The generator consists of two interconnected networks that take low-resolution over-exposed or under-exposed images as input and produce super-resolved fused images as output. Each sub-network includes an initial FEB, an SRB, and an MEFB. The FEB extracts essential features from the low-resolution inputs to support the subsequent SRB and MEFBs. With the LR over-exposed and under-exposed image inputs denoted as I_{lr}^O and I_{lr}^U , respectively, the corresponding features F_{in}^O and F_{in}^U extracted by the FEB can be obtained by the following equation:

$$\begin{aligned} F_{in}^O &= f_{FEB}(I_{lr}^O), \\ F_{in}^U &= f_{FEB}(I_{lr}^U), \end{aligned} \quad (2)$$

where f_{FEB} represents the operation of the feature extraction block. The FEB consists of two convolutional layers with parametric rectified linear unit (PReLU) activation. The first layer has 256 filters of size 3×3 , which are utilized for extracting fundamental low-resolution features. The subsequent layer utilizes 64 filters of size 1×1 to enhance and consolidate cross-channel features for more reduced feature complexity. F_{in}^O and F_{in}^U serve as important input features for the subsequent SRB and MEFBs.

With the basic features F_{in}^O and F_{in}^U as inputs, the role of the SRB is to learn more high-level features and increase image resolution. Our SRB was inspired by residual-in-residual dense block (RRDB) modules [58]. The SRB network comprises multiple dense blocks with skip connections, including one at the block's start to enhance feature transfer and prevent gradient vanishing and feature loss in deep networks, as shown in Figure 2.

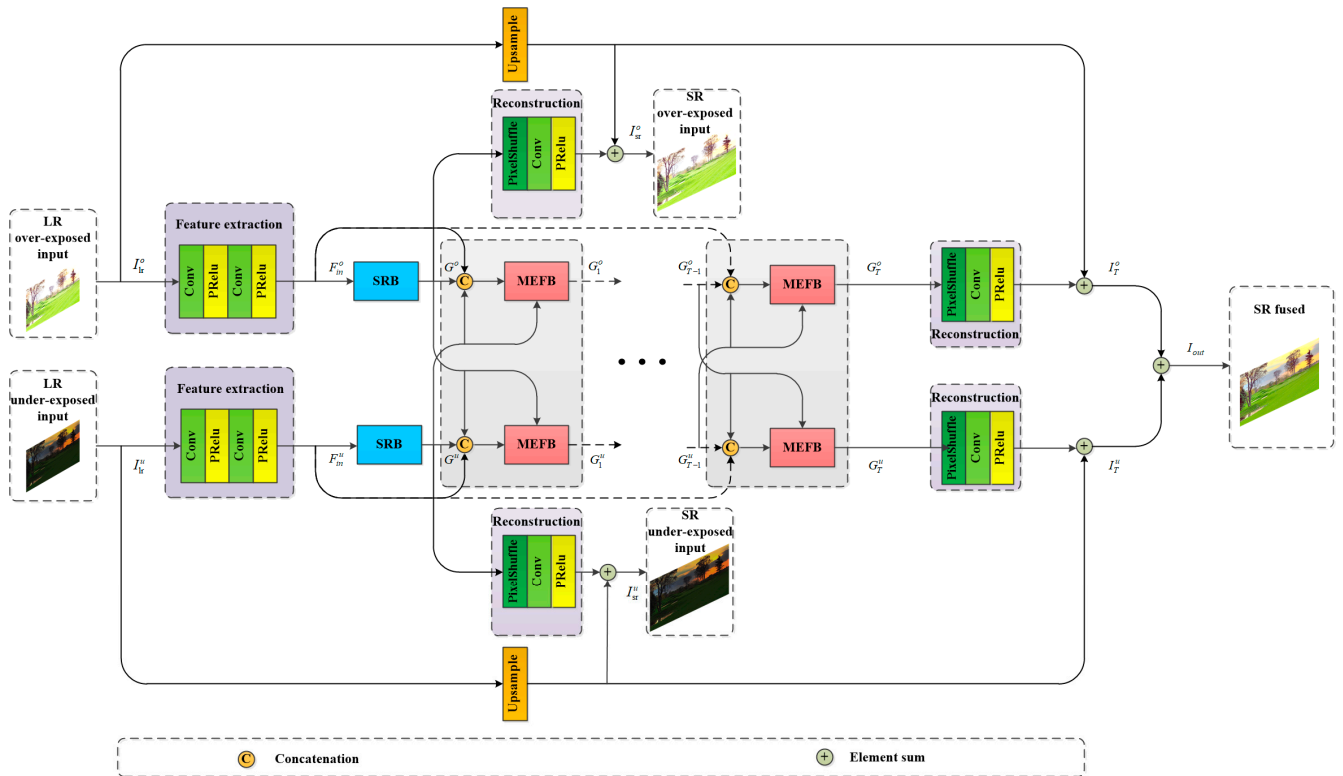


Figure 1. Network architecture of the proposed MEFSRGAN. The overall network is composed of two sub-nets with LR over-exposed and under-exposed images as inputs, respectively. Each subnet is composed of a feature extraction block (FEB), a super-resolution block (SRB), and several mu-tiple-exposure feedback blocks (MEFBs).

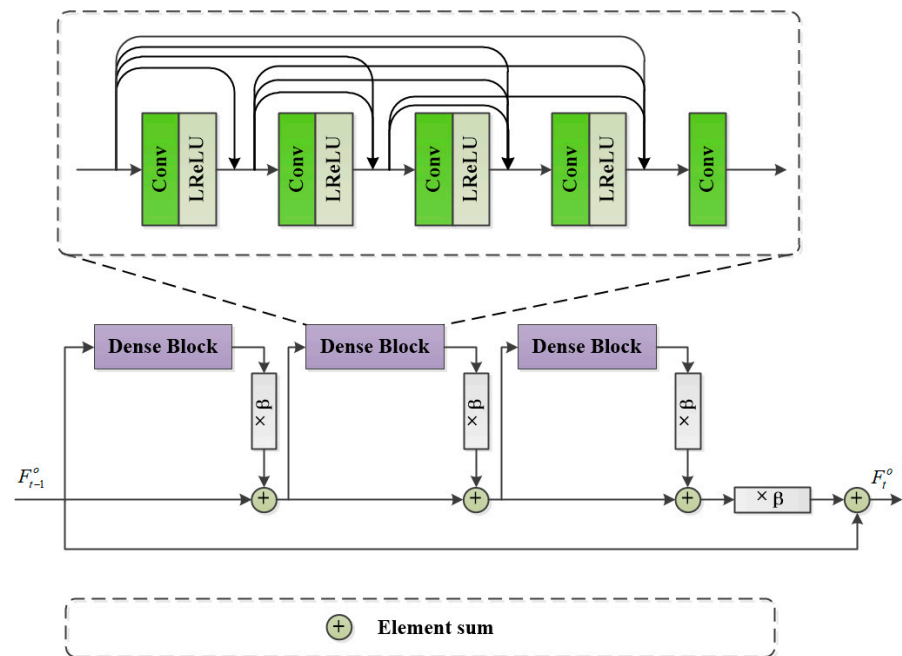


Figure 2. The super-resolution block (SRB) made up of residual-in-residual dense blocks (RRDBs).

Note that the SRB comprises multiple RRDBs. The high-level features learned by the SRB can be represented as follows:

$$\begin{aligned} G^o &= f_{\text{SRB}}(F_{\text{in}}^o), \\ G^u &= f_{\text{SRB}}(F_{\text{in}}^u), \end{aligned} \quad (3)$$

where f_{SRB} denotes the operation of the SRB and G^o and G^u represent the high-level features of the over-exposed and under-exposed images, respectively. To reconstruct the super-resolved images, a reconstruction block (REC) is utilized to map G^o and G^u to the high-resolution images. The reconstruction block features a pixel shuffle layer and a convolutional layer with PReLU. The pixel shuffle layer was chosen over deconvolution for its reduced checkerboard artifacts and greater efficiency. The original image is bilinearly upsampled and combined with the reconstructed image via a skip connection to generate super-resolution images for both over-exposed and under-exposed inputs:

$$\begin{aligned} I_{\text{sr}}^o &= f_{\text{UP}}(I_{\text{lr}}^o) + f_{\text{REC}}(G^o), \\ I_{\text{sr}}^u &= f_{\text{UP}}(I_{\text{lr}}^u) + f_{\text{REC}}(G^u), \end{aligned} \quad (4)$$

where f_{UP} represents bilinear amplification and f_{REC} represents the reconstruction operation. I_{sr}^o and I_{sr}^u are the super-resolution reconstruction results of I_{lr}^o and I_{lr}^u , respectively, without multi-exposure features. By constraining the loss functions of I_{lr}^o and I_{lr}^u , the SRB module ensures effective operation and offers reliable high-level features for subsequent MEFBs.

The MEFBs are the core of our proposed MEFSR-GAN, designed to achieve super resolution and image fusion simultaneously through interconnections (see Figure 1). Unlike the FEB and SRB, multiple interconnected MEFBs are employed, where the output of the $(T - 1)$ -th MEFB serves as the input for the T -th MEFB, along with two additional inputs. For example, given I_{lr}^o , the output of the T -th MEFB is represented as follows:

$$G_T^o = f_{\text{MEFB}}(F_{\text{in}}^o, G_{T-1}^o, G_{T-1}^u), \quad (5)$$

where F_{in}^o represents the basic feature extracted by MEFB, and G_{T-1}^o and G_{T-1}^u come from the previous over-exposed and under-exposed MEFB, respectively. In (5), the first two inputs play a significant role in enhancing super-resolution performance while the last input enhances fusion effects. Similarly, for an MEFB with I_{lr}^u as input, the output of the T -th MEFB can be expressed as follows:

$$G_T^u = f_{\text{MEFB}}(F_{\text{in}}^u, G_{T-1}^u, G_{T-1}^o), \quad (6)$$

After each MEFB, we can reconstruct a fused SR image by the following equation:

$$\begin{aligned} I_T^o &= f_{\text{UP}}(I_{\text{lr}}^o) + f_{\text{REC}}(G_T^o), \\ I_T^u &= f_{\text{UP}}(I_{\text{lr}}^u) + f_{\text{REC}}(G_T^u), \end{aligned} \quad (7)$$

where f_{UP} represents an upsampling operation and f_{REC} represents a reconstruction operation. Both I_T^o and I_T^u are super-resolution images with high dynamic range. Since G_T^o and G_T^u are generated by MEFBs from over-exposed and under-exposed images, both I_T^o and I_T^u have high dynamic range features. After processing, the final high-dynamic-range super-resolution image (HDR-SR) is obtained using the following formula:

$$I_{\text{out}} = \eta_o I_T^o + \eta_u I_T^u, \quad (8)$$

In this study, we set both η_o and η_u to 0.5 as weighted parameters. Although not the final images, I_T^o and I_T^u combine their feature information post-MEFB processing. These images already approach the high-dynamic-range ground truth image within the limitations

of the loss function. Experimental results demonstrate that I_T^o and I_T^u exhibit good dynamic range and high resolution. Through the application of the constraint fusion described in Formula (8), superior fusion results can be attained.

3.2. Multi-Exposure Feedback Block (MEFB)

The multiple exposure feedback block (MEFB) is a fundamental component of MEFSR-GAN. Prior studies [59,60] highlight the importance of feedback mechanisms in image restoration and super resolution. This paper introduces a multi-exposure feedback mechanism that enhances both tasks simultaneously. The architecture of the MEFB is shown in Figure 3. Multiple MEFBs are used consecutively throughout the network; we focus on the t -th MEFB in the over-exposure subnetwork to detail its structure and interactions.

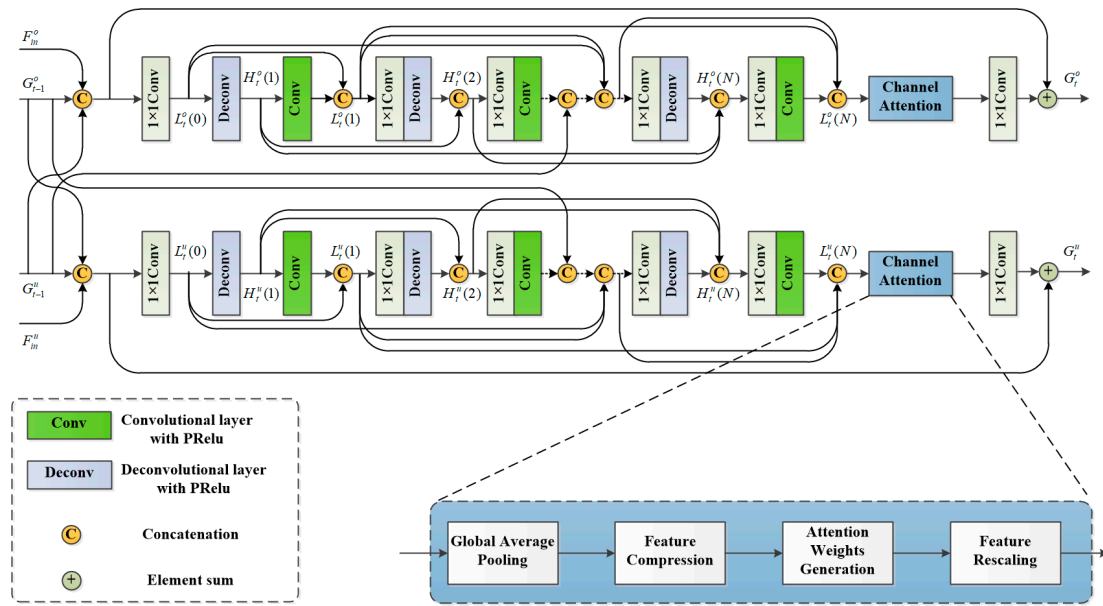


Figure 3. Architecture of the multiple exposure feedback block (MEFB). The upper sub-network of the T -th MEFB accepts F_{in}^o , G_{T-1}^o and G_{T-1}^u as inputs, and outputs G_T^o , while the lower sub-network of the T -th MEFB accepts F_{in}^u , G_{T-1}^u and G_{T-1}^o as inputs, and outputs G_T^u .

In the upper layer of the over-exposure subnetwork, the t -th MEFB receives three inputs: a basic feature F_{in}^o extracted by the FEB, and two feedback features G_{t-1}^o and G_{t-1}^u from the previous MEFB. These latter two feedback features have different roles. Specifically, G_{t-1}^o is a feedback feature from the same sub-network and is primarily aimed at correcting and boosting the basic feature F_{in}^o to enhance SR performance. On the other hand, G_{t-1}^u represents feedback from the other sub-network and its primary function is to incorporate feature information from under-exposed images to enhance fusion performance. Taken together, these three feature inputs are first concatenated and then fused by a set of 1×1 filters as follows:

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

$$L_t^o(0) = f_{conv}(L_t^o), \quad (9)$$

where $L_t^o(0)$ is the refined feature based on the three inputs, while f_{conv} represents a set of 1×1 filters. The f_{cat} operation combines the three inputs in the feature dimension. This type of connection is commonly utilized for feature fusion, and it involves merging data from different feature sets into a larger feature set to enhance the input information for further processing by the model. Essentially, this approach aims to expand the feature space to enable the model to receive and process more information. Subsequently, a series of projection groups repeatedly perform upscaling and downscaling operations with $L_t^o(0)$ as input to extract more effective high-level features. Within each projection group,

upsampling is initially carried out through a deconvolution layer to obtain HR feature maps; this is followed by downsampling through a convolutional layer to generate LR feature maps. Building upon previous research [59,60], we utilize dense connections to incorporate all previously extracted features for both upsampling and downsampling. Let $L_t^o(n)$ and $H_t^o(n)$, respectively, represent the LR and HR feature maps extracted in the n -th projection group within the t -th MEFB. The HR feature map $H_t^o(n)$ can be obtained through the following process:

$$H_t^o(n) = f_{deconv}([L_t^o(0), L_t^o(1), L_t^o(2), \dots, L_t^o(n-1)]), \quad (10)$$

where f_{deconv} represents the deconvolution operation in the n -th projection group. It is evident that all previous LR feature maps are combined to produce HR feature maps. Likewise, the LR feature map $L_t^o(n)$ in the n -th projection group is created by combining all previous HR feature maps:

$$L_t^o(n) = f_{conv}([H_t^o(0), H_t^o(1), H_t^o(2), \dots, H_t^o(n-1)]), \quad (11)$$

where f_{conv} represents the convolution operation in the n -th projection group. As previously discussed, the feedback feature G_{t-1}^u from the under-exposed sub-network is essential for improving fusion performance. However, it may be noticed that as the number of projection groups increases, the influence of G_{t-1}^u decreases, and this results in less effective fusion outcomes. This decrease in influence is mainly due to the gradual weakening or disappearance of the feature memory of G_{t-1}^u with longer projection groups. To strengthen the influence of G_{t-1}^u , we not only use it as initial input for the MEFBs, but also integrate it into the middle layer of the projection group through skip connections to revive the memory. Assuming a total of N projection groups, the reactivation of fusion features by G_{t-1}^u is completed at the position of the M -th projection group, where M can be expressed as follows:

$$M = \text{round}\left(\frac{N}{2}\right), \quad (12)$$

where $\text{round}(\cdot)$ denotes a round operation. The skip connection of G_{t-1}^u in the M -th projection group can be expressed as follows:

$$L_t^o(M)^* = f_{cat}(L_t^o(M), G_{t-1}^u), \quad (13)$$

where $L_t^o(M)$ represents the LR feature map in the n -th projection group and f_{cat} denotes a concatenation operation. After N projection groups, the LR feature maps are collected and then passed to the channel attention module.

In order to assess and adjust the significance of various channel features, we have integrated a channel attention mechanism module into the MEFB. This module comprises four layers: global average pooling, feature compression, attention weights generation, and feature rescaling. Our attention mechanism effectively evaluates and adjusts the importance of different channel features and thus improves the network's sensitivity to key information. This process is crucial for enhancing MEFSR-GAN's performance and generalizability.

After channel attention, the LR feature maps are fused together using a set of 1×1 filters. The input features L_t^o are then added to the N -th projection group via residual connections to form the output of the t -th MEFB. The formula for this process is as follows:

$$G_t^o = f_{conv}([L_t^o(0), L_t^o(1), \dots, L_t^o(M)^*, \dots, L_t^o(N)]) + L_t^o, \quad (14)$$

where f_{conv} denotes the convolution using a set of 1×1 filters. The output of the t -th MEFB G_t^o is passed to the $(t+1)$ -th MEFB as input. The same feature learning process as for the t -th MEFB is repeated. Following the reconstruction operation, the new features from G_T^o are

connected with the upsampled features of the original image to produce super-resolution images with a high dynamic range, which are denoted as I_T^o .

The network structure of the MEFB over-exposed sub-network is discussed in the preceding section. The MEFB under-exposed sub-network mirrors the same architecture as the upper sub-network, as illustrated in Figure 3, and thus is not further elaborated on in this paper.

3.3. Discriminator Network

This study introduces a U-Net discriminator with spectral normalization (SN) to address image degradation (see Figure 4). The discriminator provides precise feedback on both style and textural details. By adding skip connections to a VGG-style discriminator, the U-Net structure enhances pixel-level feedback but increases training instability due to its complexity. To stabilize training and reduce over-sharpening artifacts, we apply spectral normalization. These adjustments effectively balance the enhancement of local details and the reduction of artifacts in MEFSR-GAN training.

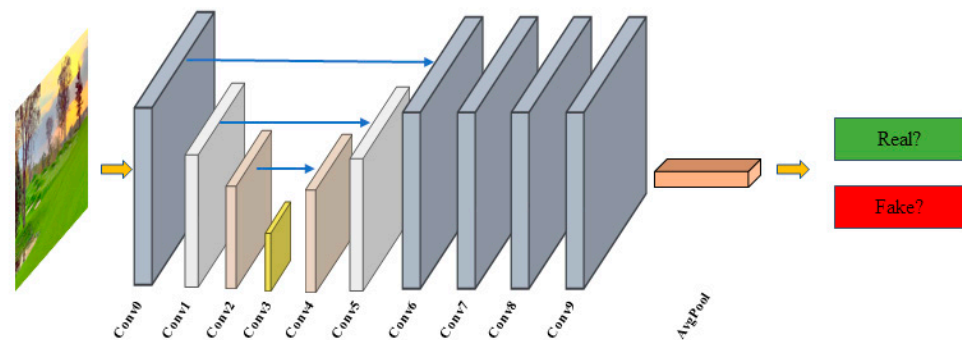


Figure 4. The U-Net discriminator network with spectral normalization.

3.4. Loss Function

3.4.1. Adversarial Loss

We introduce an enhanced discriminator called the relativistic average discriminator (RaD) [61] to replace the standard discriminator in SRGAN. In SRGAN, the standard discriminator is typically denoted as $D(x) = \sigma(C(x))$, where σ represents a sigmoid function, and $C(x)$ is the discriminator's output without any transformation. Conversely, the RaD formula is expressed as $D_{Ra}(x_r, x_f) = \sigma(C(x_r) - \mathbb{E}_{x_f}[C(x_f)])$, where $\mathbb{E}_{x_f}[\cdot]$ represents the average computation for all synthetic data in a mini-batch. Subsequently, the discriminator loss is derived as follows:

$$L_{adv}(D(x_t), y_t) = \begin{aligned} & -\mathbb{E}_{x_r}[\log(D_{Ra}(y_t, G(x_t)))] \\ & -\mathbb{E}_{x_f}[\log(1 - D_{Ra}(G(x_i), y_i))], \end{aligned} \quad (15)$$

The adversarial loss for the generator takes a symmetrical form:

$$L_{adv}(G(x_i), y_i) = \begin{aligned} & -\mathbb{E}_{x_r}[\log(1 - D_{Ra}(y_i, G(x_i)))] \\ & -\mathbb{E}_{x_f}[\log(D_{Ra}(G(x_i), y_i))], \end{aligned} \quad (16)$$

where y_i represents the target image and $G(x_i)$ represents the SR image. This loss function encourages the generator to produce images that are challenging for the discriminator to differentiate, which ultimately enhances the realism of the generated images.

3.4.2. Content Loss

In super-resolution (SR) tasks, content loss is critical for measuring similarity between the generated and real images, typically using mean square error (L2 loss) or absolute error (L1 loss). L1 loss is favored for its ability to reduce sensitivity to outliers, promoting smooth images with sharp edges and better preserving intricate details and textures. Compared

with L2 loss, L1 loss excels in handling high-frequency details by avoiding excessive smoothing, making it widely used in image super-resolution reconstruction. Specifically, $L_{content}$ can be represented as follows:

$$L_{content}(G(x_i), y_i) = \mathbb{E}_{x_i}[\|G(x_i) - y_i\|_1], \quad (17)$$

where the notation $\mathbb{E}_{x_i}[\cdot]$ denotes the expectation for all pixels x_i , $G(x_i)$ represents the SR image, and y_i represents the target image.

3.4.3. Perceptual Loss

The perceptual loss evaluates the perceptual quality difference between the generated and target images. Using the first 20 layers of the pre-trained VGG-19 model [62], we extract image features and apply L1 loss to measure feature differences. The formula for perceptual loss is as follows:

$$L_{perceptual}(G(x_i), y_i) = \mathbb{E}_{x_i}[\|VGG(G(x_i)) - VGG(y_i)\|_1], \quad (18)$$

Using VGG network features to compute loss provides a comprehensive assessment of perceptual quality beyond pixel-level differences, making it particularly effective for image super-resolution and fusion tasks.

3.4.4. Structural Similarity Index Measure Loss

SSIM loss is widely used in image processing, especially for image quality assessment and reconstruction [63]. Unlike pixel-based losses like L1 and L2, SSIM loss emphasizes perceptual quality by evaluating contrast, brightness, and structure, aligning more closely with human visual perception. The SSIM loss function is defined as follows:

$$L_{ssim}(y, \hat{y}) = 1 - \frac{(2\mu_y\mu_{\hat{y}} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2)}, \quad (19)$$

where y and \hat{y} represent the SR image and the target image, respectively, μ_y and $\mu_{\hat{y}}$ are the average brightnesses of those images, σ_y^2 and $\sigma_{\hat{y}}^2$ are the variances of the images, $\sigma_{y\hat{y}}$ is the covariance of y and \hat{y} , and C_1 and C_2 are constants introduced to stabilize the division operation.

SSIM loss quantifies image similarity by prioritizing structural characteristics, enabling optimization to focus on preserving image integrity rather than just pixel matching. In tasks like super-resolution and denoising, SSIM loss enhances the naturalness and accuracy of reconstructed images, making it a valuable tool in image processing.

3.4.5. Mixed Loss Function

Since we aim to achieve super-resolution and image fusion simultaneously, we impose hierarchical loss constraints to guarantee efficient network training. The total loss function of our MEFSR-GAN is defined as follows:

$$\begin{aligned} L_{total} = & \lambda_t(L_{ssim}(I_{sr}^o, I_{gt}^o) + L_{ssim}(I_{sr}^u, I_{gt}^u)) + \lambda_a(L_{adv}(I_T^o, I_{gt}) + L_{adv}(I_T^u, I_{gt})) \\ & + \lambda_p(L_{perceptual}((I_T^o, I_{gt}) + L_{perceptual}(I_T^u, I_{gt})) + \lambda_c((L_{content}((I_T^o, I_{gt}) \\ & + (L_{content}(I_T^u, I_{gt})) + \lambda_s(L_{ssim}((I_T^o, I_{gt}) + L_{ssim}(I_T^u, I_{gt})) \end{aligned} \quad (20)$$

where I_{gt}^o and I_{gt}^u are the ground truth high-resolution (HR) over-exposed and under-exposed images, respectively, and I_{gt} is the ground truth HR image with high dynamic range, which is our final target. I_T^o and I_T^u represent the super-resolution (SR) over-exposed and under-exposed images, respectively, that have been processed by the MEF network. λ_t , λ_a , λ_p , λ_c , and λ_s are the weights assigned to each loss. The loss function in Equation (20) comprises two parts: the first ensures SRB effectiveness, optimizing super-resolution performance, while the second maintains MEB functionality, enhancing both SR and MEF

performance. The first part also underpins the losses in the second. The network is trained end-to-end by minimizing this loss function.

4. Results

The performance of the proposed MEFSR-GAN is evaluated in this section. The experimental setup is detailed in Section 4.1, while Sections 4.2 and 4.3 present the quantitative and qualitative comparison results with other state-of-the-art methods. Ablation studies are described in Section 4.4.

4.1. Experimental Setup

4.1.1. Dataset

The training data were sourced from the SICE dataset [64], containing images with various exposure levels. To address extreme exposure image fusion and super-resolution, we specifically chose pairs of highly over-exposed and under-exposed images from the dataset for training. Figure 3 shows examples of these images, encompassing diverse scenes containing people, natural landscapes, and man-made structures. Notably, the under-exposed images appear very dark while the over-exposed images are excessively bright; both conceal significant detail. Through the MEFSR network, we were able to merge these concealed details and enhance image resolution. During network training, the SICE dataset provided real fusion images for HDR reference. From this dataset, we randomly selected 420 pairs of over-exposed and under-exposed images: 300 pairs for training and 100 pairs for testing. In addition to training and testing, we also used 20 pairs of images for validation. Apart from the test image pairs sourced from the SICE dataset, the PQA-MEF dataset [65] was also utilized for testing purposes. During training, we utilized data enhancement technology to further expand the training data.

4.1.2. Training Details

To generate LR training images, we applied bicubic downsampling to HR images that were over-exposed and under-exposed, using MATLAB's bicubic kernel with $2\times$ and $4\times$ downsampling factors, respectively. The mini-batch size was set to 32, and the spatial size of the cropped LR patch was 40×40 . Our findings suggested that training a deeper network benefited from a larger patch size since it enhanced the receptive field for capturing more semantic information. However, this came with the drawback of longer training times and increased consumption of computing resources. Each subnetwork was composed of three MEFBs, with each MEFB containing six projection groups.

The generator was trained using the loss function defined in Equation (3) with $\lambda_t = 1$, $\lambda_a = 0.1$, $\lambda_p = 0.01$, $\lambda_c = 0.1$, and $\lambda_s = 1$. The generator's learning rate was set to 1×10^{-4} , while the discriminator's learning rate was set to 1×10^{-4} . Both the generator and discriminator had their learning rates reduced by half at iterations 2500 and 4500. Optimization was performed using Adam with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The output weights η_o and η_u were both set to 0.5. Finally, the generator was trained for 1×10^5 epochs. The generator and discriminator were iteratively updated until the model reached convergence.

4.1.3. Comparison Methods

Our study aimed to complete exposure fusion and image super-resolution simultaneously. Currently, only CF-Net has achieved this through a multi-task network. We identified several state-of-the-art methods for solving SR or MEF problems and combined them to develop a method able to address both SR and MEF. We also considered several SR methods (i.e., EDSR [39], SRFBN [60], SWINIR [42], and RND [66]) and MEF methods (i.e., IFCNN [53], MEF-Net [67], U2Fusion [57], and Fast SPD-MEF [68]). To compare these methods, a new approach was developed, randomly combining different SR and MEF methods and altering the order of application: i.e., either performing SR followed by MEF (SR+MEF) or MEF followed by SR (MEF+SR). All deep learning-based models

were retrained using the same training dataset as ours for fair comparison. The training process involved training the first task, then the second task, and finally fine-tuning the entire network to achieve optimal results.

4.2. Quantitative Comparison Results

In this study, we assessed the effectiveness of our method using three evaluation metrics: PSNR, SSIM [63], and MEF-SSIM [68]. Higher values of these indicators indicate better performance in image super-resolution and fusion. Results obtained with the SICE dataset included evaluation of these three indicators. For results obtained with the PQA-MEF dataset, lacking real data, we evaluated only the MEF-SSIM.

Tables 1 and 2 compare the proposed method with other state-of-the-art methods. We focus on comparisons of $2\times$ SR and MEF using the SICE and PQA-MEF datasets. The tables display experimental results derived from executing SR and MEF in both orders.

Table 1. Comparisons with other state-of-the-art methods in terms of PSNR, SSIM, and MEF-SSIM using the SICE dataset for $2\times$ upscaling. The best results are in bold and the second-best results are underlined.

| SR + MEF | | | | | | | | | | | | |
|----------|----------------------|-------|----------|---------------------|-------|----------|-------------------------|-------|----------|----------|-------|----------|
| Methods | IFCNN | | | MEF-Net | | | Fast SPD | | | U2Fusion | | |
| | PSNR | SSIM | MEF-SSIM | PSNR | SSIM | MEF-SSIM | PSNR | SSIM | MEF-SSIM | PSNR | SSIM | MEF-SSIM |
| EDSR | 21.003 | 0.812 | 0.808 | 15.127 | 0.765 | 0.794 | 17.088 | 0.757 | 0.829 | 16.119 | 0.723 | 0.810 |
| SRFBN | 20.949 | 0.808 | 0.803 | 15.106 | 0.760 | 0.786 | 17.052 | 0.752 | 0.819 | 16.098 | 0.718 | 0.819 |
| RDN | 20.910 | 0.796 | 0.790 | 15.093 | 0.753 | 0.774 | 17.063 | 0.743 | 0.798 | 16.088 | 0.714 | 0.774 |
| SWINIR | 21.034 | 0.816 | 0.811 | 15.153 | 0.770 | 0.799 | 17.149 | 0.763 | 0.840 | 16.131 | 0.727 | 0.816 |
| MEF + SR | | | | | | | | | | | | |
| Methods | IFCNN | | | MEF-Net | | | Fast SPD | | | U2Fusion | | |
| | PSNR | SSIM | MEF-SSIM | PSNR | SSIM | MEF-SSIM | PSNR | SSIM | MEF-SSIM | PSNR | SSIM | MEF-SSIM |
| EDSR | 20.251 | 0.777 | 0.756 | 15.106 | 0.763 | 0.787 | 16.908 | 0.739 | 0.796 | 16.031 | 0.686 | 0.757 |
| SRFBN | 20.286 | 0.776 | 0.754 | 15.090 | 0.759 | 0.781 | 16.935 | 0.739 | 0.796 | 16.022 | 0.684 | 0.754 |
| RDN | 20.347 | 0.774 | 0.757 | 15.071 | 0.752 | 0.771 | 16.962 | 0.736 | 0.793 | 16.028 | 0.683 | 0.751 |
| SWINIR | 20.082 | 0.772 | 0.749 | 15.118 | 0.766 | 0.790 | 16.867 | 0.739 | 0.791 | 16.035 | 0.688 | 0.759 |
| CF-Net | PSNR = <u>22.669</u> | | | SSIM = <u>0.857</u> | | | MEF-SSIM = <u>0.848</u> | | | | | |
| Ours | PSNR = 24.821 | | | SSIM = 0.896 | | | MEF-SSIM = 0.855 | | | | | |

Table 2. Comparisons with other state-of-the-art methods in terms of MEF-SSIM using the PQA-MEF dataset for $2\times$ upscaling. The best results are in bold and the second-best results are underlined.

| SR + MEF | | | | |
|----------|-------|---------|----------|--------------|
| Methods | IFCNN | MEF-Net | Fast SPD | U2Fusion |
| EDSR | 0.774 | 0.801 | 0.869 | 0.867 |
| SRFBN | 0.782 | 0.806 | 0.870 | 0.870 |
| RDN | 0.787 | 0.810 | 0.861 | <u>0.875</u> |
| SWINIR | 0.762 | 0.796 | 0.864 | 0.861 |
| MEF + SR | | | | |
| Methods | IFCNN | MEF-Net | Fast SPD | U2Fusion |
| EDSR | 0.744 | 0.800 | 0.842 | <u>0.875</u> |
| SRFBN | 0.752 | 0.804 | 0.847 | 0.877 |
| RDN | 0.761 | 0.808 | 0.849 | <u>0.875</u> |
| SWINIR | 0.732 | 0.794 | 0.828 | 0.874 |
| CF-Net | | | | 0.851 |
| Ours | | | | 0.843 |

The results in Table 1 demonstrate that our method outperformed all others on the SICE dataset. Specifically, we achieved a 2.152 dB higher PSNR, 0.039 higher SSIM, and 0.007 higher MEF-SSIM compared with the second-best method.

As shown in Table 2, for the PQA-MEF dataset, only MEF-SSIM was used as the evaluation metric due to the absence of a ground truth image. The reference input consisted of sequences of over-exposed and under-exposed images. The wide dynamic range of extremely exposed images led to missing details. Note that MEF-SSIM may not allow us to accurately assess the color reproduction quality of an image. Consequently, even if an image lacks good detail, structure, or accurate color reproduction, it may still yield a high MEF-SSIM result. However, such images may be perceived as poor quality by the human eye. The results in Table 2 further support this observation. Therefore, MEF-SSIM alone was not fully able to evaluate the quality of extremely exposed color images in this experiment. For a more detailed comparison of image characteristics, please refer to the subsequent subsection on qualitative comparisons.

Tables 3 and 4 present the $4\times$ upscaling results obtained with the SICE and PQA-MEF datasets, respectively. Using the SICE dataset, our method obtained a 0.963 dB higher PSNR, 0.007 higher SSIM, and 0.058 higher MEF-SSIM than the second-best method. For the PQA-MEF dataset, the test result was similar to $\times 2$. The absence of ground truth images necessitated the use of over-exposed and under-exposed grayscale image sequences as inputs for calculating MEF-SSIM. Consequently, there was a notable deficiency in color details and the broad dynamic range that the ground truth images possessed. As a result, while the MEF-SSIM scores from the U2Fusion + RDN compared method were notably high when using the PQA-MEF dataset, the final image still lagged behind that produced by our proposed method in terms of color accuracy and resolution. MEF-SSIM is not particularly effective at capturing halos and may even exhibit a preference for this artifact [69]. Therefore, the performance of our method may not have been fully demonstrated by the MEF-SSIM values. For more detailed results and analysis, please refer to the upcoming subsection on qualitative comparisons.

Table 3. Comparisons with other state-of-the-art methods in terms of PSNR, SSIM, and MEF-SSIM using the SICE dataset for $4\times$ upscaling. The best results are in bold and the second-best results are underlined.

| SR + MEF | | | | | | | | | | | | |
|----------|----------------------|-------|----------|---------------------|-------|----------|-------------------------|-------|----------|----------|-------|--------------|
| Methods | IFCNN | | | MEF-Net | | | Fast SPD | | | U2Fusion | | |
| | PSNR | SSIM | MEF-SSIM | PSNR | SSIM | MEF-SSIM | PSNR | SSIM | MEF-SSIM | PSNR | SSIM | MEF-SSIM |
| EDSR | 19.813 | 0.662 | 0.643 | 14.727 | 0.627 | 0.618 | 16.559 | 0.616 | 0.627 | 15.562 | 0.587 | 0.633 |
| SRFBN | 19.814 | 0.673 | 0.657 | 14.722 | 0.630 | 0.623 | 16.568 | 0.623 | 0.643 | 15.579 | 0.588 | 0.636 |
| RDN | 19.632 | 0.620 | 0.595 | 14.270 | 0.507 | 0.526 | 15.972 | 0.512 | 0.543 | 15.418 | 0.526 | 0.584 |
| SWINIR | 20.054 | 0.711 | 0.702 | 14.828 | 0.665 | 0.668 | 16.726 | 0.657 | 0.701 | 15.689 | 0.623 | <u>0.685</u> |
| MEF + SR | | | | | | | | | | | | |
| Methods | IFCNN | | | MEF-Net | | | Fast SPD | | | U2Fusion | | |
| | PSNR | SSIM | MEF-SSIM | PSNR | SSIM | MEF-SSIM | PSNR | SSIM | MEF-SSIM | PSNR | SSIM | MEF-SSIM |
| EDSR | 19.020 | 0.646 | 0.616 | 14.718 | 0.626 | 0.615 | 16.448 | 0.601 | 0.616 | 15.532 | 0.555 | 0.590 |
| SRFBN | 18.990 | 0.651 | 0.619 | 14.705 | 0.629 | 0.619 | 16.454 | 0.606 | 0.621 | 15.540 | 0.555 | 0.592 |
| RDN | 19.050 | 0.626 | 0.603 | 14.677 | 0.604 | 0.596 | 16.420 | 0.580 | 0.595 | 15.550 | 0.546 | 0.582 |
| SWINIR | 18.525 | 0.651 | 0.614 | 14.760 | 0.656 | 0.650 | 16.254 | 0.612 | 0.623 | 15.584 | 0.568 | 0.605 |
| CF-Net | PSNR = <u>20.965</u> | | | SSIM = <u>0.722</u> | | | MEF-SSIM = 0.678 | | | | | |
| Ours | PSNR = 21.928 | | | SSIM = 0.729 | | | MEF-SSIM = 0.743 | | | | | |

Table 4. Comparisons with other state-of-the-art methods in terms of MEF-SSIM using the PQA-MEF dataset for 4× upscaling. The best results are in bold and the second-best results are underlined.

| SR + MEF | | | | |
|----------|-------|---------|----------|--------------|
| Methods | IFCNN | MEF-Net | Fast SPD | U2Fusion |
| EDSR | 0.788 | 0.807 | 0.835 | 0.850 |
| SRFBN | 0.791 | 0.806 | 0.845 | 0.843 |
| RDN | 0.791 | 0.823 | 0.823 | <u>0.869</u> |
| SWINIR | 0.739 | 0.776 | 0.820 | 0.809 |
| MEF + SR | | | | |
| Methods | IFCNN | MEF-Net | Fast SPD | U2Fusion |
| EDSR | 0.749 | 0.804 | 0.815 | 0.848 |
| SRFBN | 0.750 | 0.803 | 0.818 | 0.845 |
| RDN | 0.767 | 0.822 | 0.825 | 0.875 |
| SWINIR | 0.693 | 0.772 | 0.774 | 0.846 |
| CF-Net | | | 0.766 | |
| Ours | | | 0.748 | |

4.3. Qualitative Comparison Results

In this subsection, we qualitatively compare our method against others. As shown in Figure 5, when using the compared methods, incorrect colors were visible in the skies in the images (c,d) and (g,h); moreover, an incorrect halo was present around the two peaks seen in image (k). Meanwhile, our method effectively maintained global image contrast and accurately restored image detail and color information.

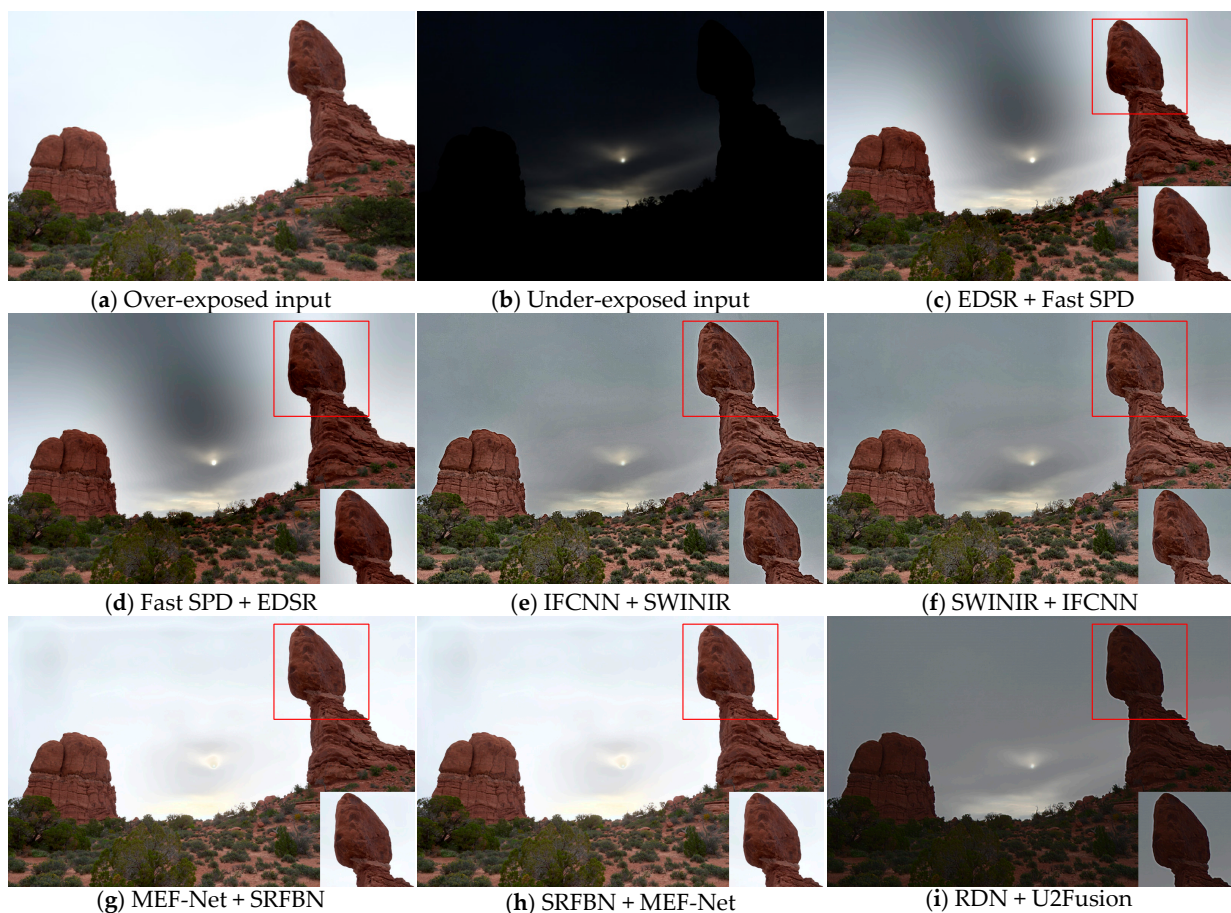


Figure 5. Cont.

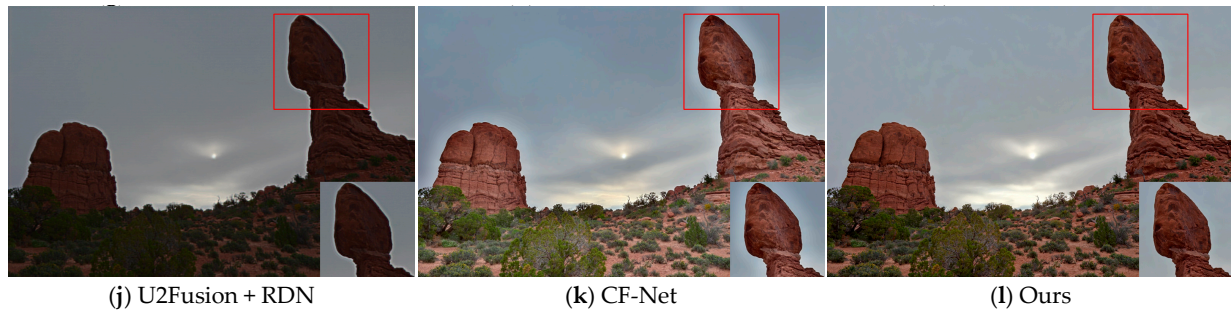


Figure 5. Visual comparison of super-resolution and exposure fusion results from experiments on the SICE dataset with $2\times$ upscaling: (a,b) over-exposed and under-exposed input images, respectively; (c–k) results obtained by the compared methods; and (l) result obtained by our method.

Figure 6 visualizes the results obtained by our method and the other methods using the PQA-MEF dataset. Images obtained via methods (c,d) and (g,h) all exhibited numerous erroneous black features in regions with higher brightness. In contrast, method (k) led to distorted and blurred letters on the hot air balloon. Methods (i,j) produced darker overall brightness and serious color distortion. Our method stands out by preserving higher details and contrast.

Figure 7 presents the results obtained on the SICE dataset with $4\times$ upscaling. In the results generated by methods (c,d) and (g,h), there were noticeable overly bright or overly dark artifacts present in the sky. Meanwhile, methods (i,j) exhibited an overall dull color palette with a lack of detail. Additionally, methods (e,f) failed to accurately restore resolution and texture details. In contrast, our proposed method effectively restored color information and thus provided a higher dynamic range and resolution. Examining locally enlarged features, it is evident that our method outperforms others in terms of resolution enhancement and fusion.



Figure 6. Cont.



Figure 6. Visual comparison of super-resolution and exposure fusion results from experiments on the PQA-MEF dataset with $2\times$ upscaling: (a,b) over-exposed and under-exposed input images, respectively; (c–k) results obtained by the compared methods; and (l) result obtained by our method.

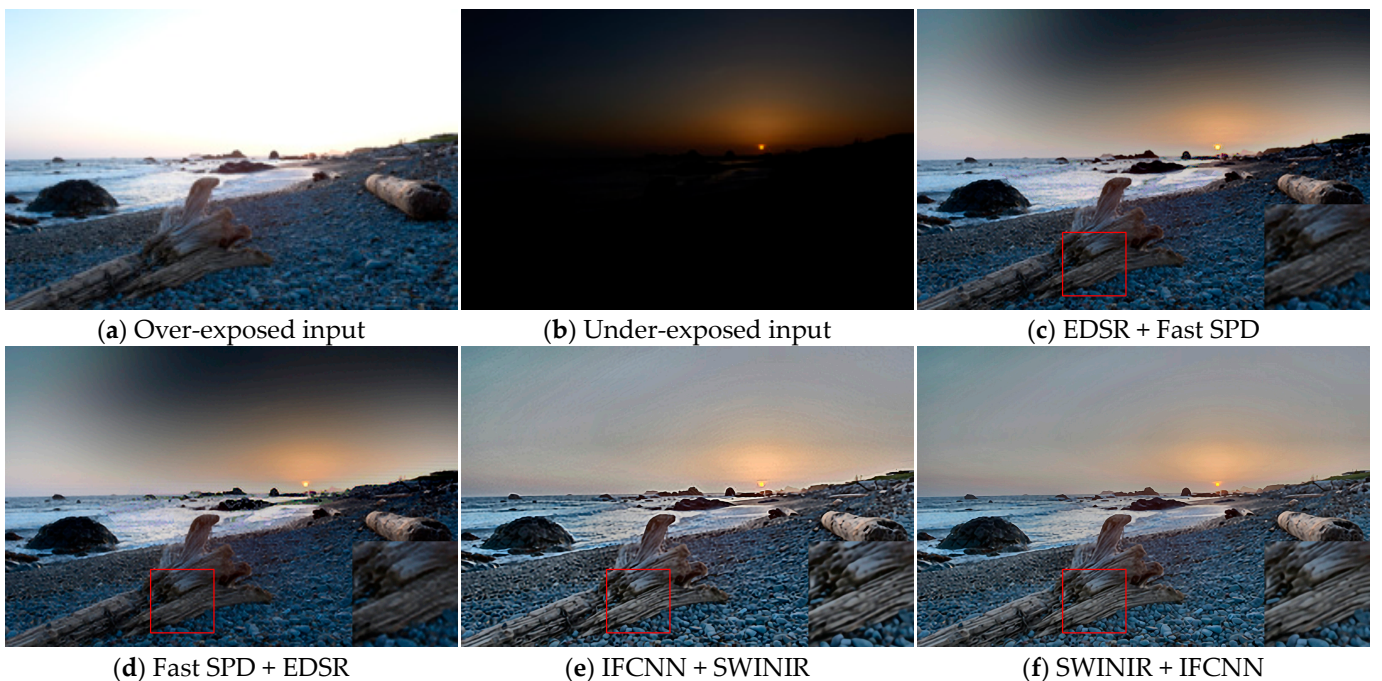


Figure 7. Cont.



Figure 7. Visual comparison of super-resolution and exposure fusion results from experiments on the SICE dataset with $4\times$ upscaling: (a,b) over-exposed and under-exposed input images, respectively; (c–k) results obtained by the compared methods; and (l) result obtained by our method.

Figure 8 displays results obtained using the PQA-MEF dataset with $4\times$ upscaling, specific to images of candles. Images obtained with methods (c,d) exhibited a significant number of artifacts, while those using methods (e,f) and (i,j) appeared dim in color, lacked highly dynamic features, and had low resolution. In the result produced by method (k), erroneous black features were visible on the desktop. In comparison, our proposed method accurately restored the overall color information of the image as well as correctly capturing the details of the candle flame's core characteristics and highly dynamic brightness features.

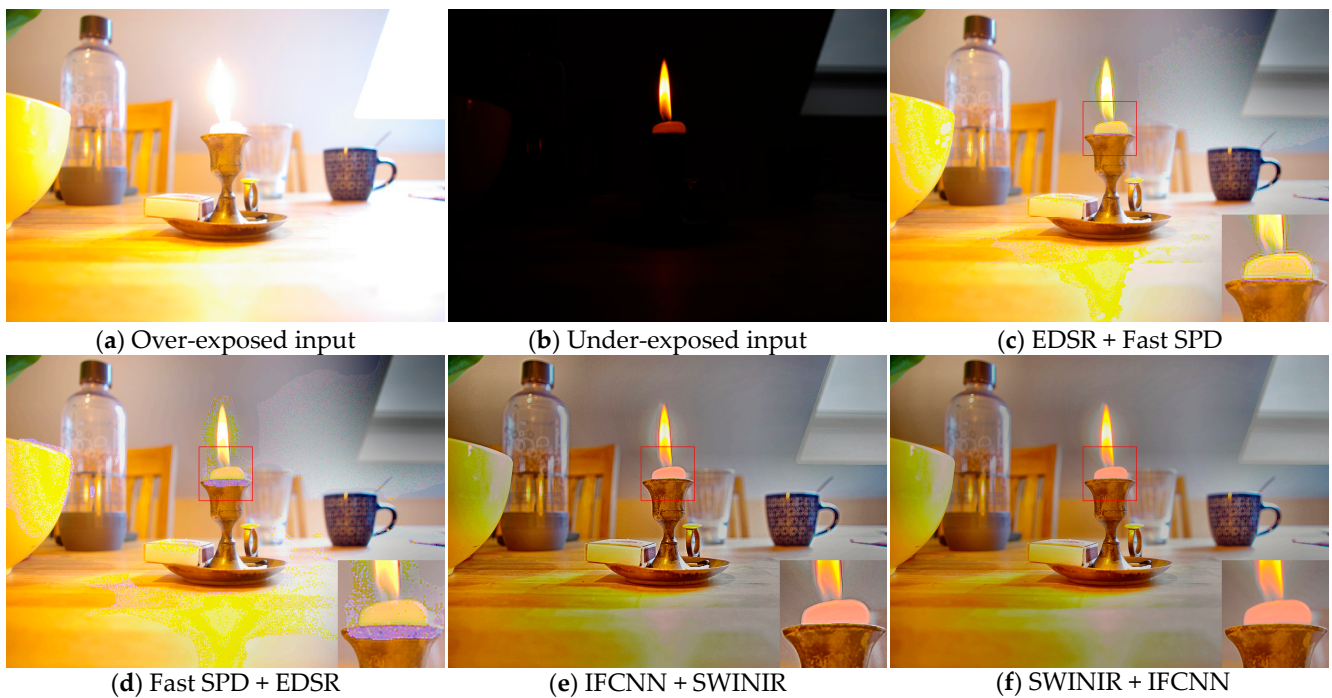


Figure 8. Cont.

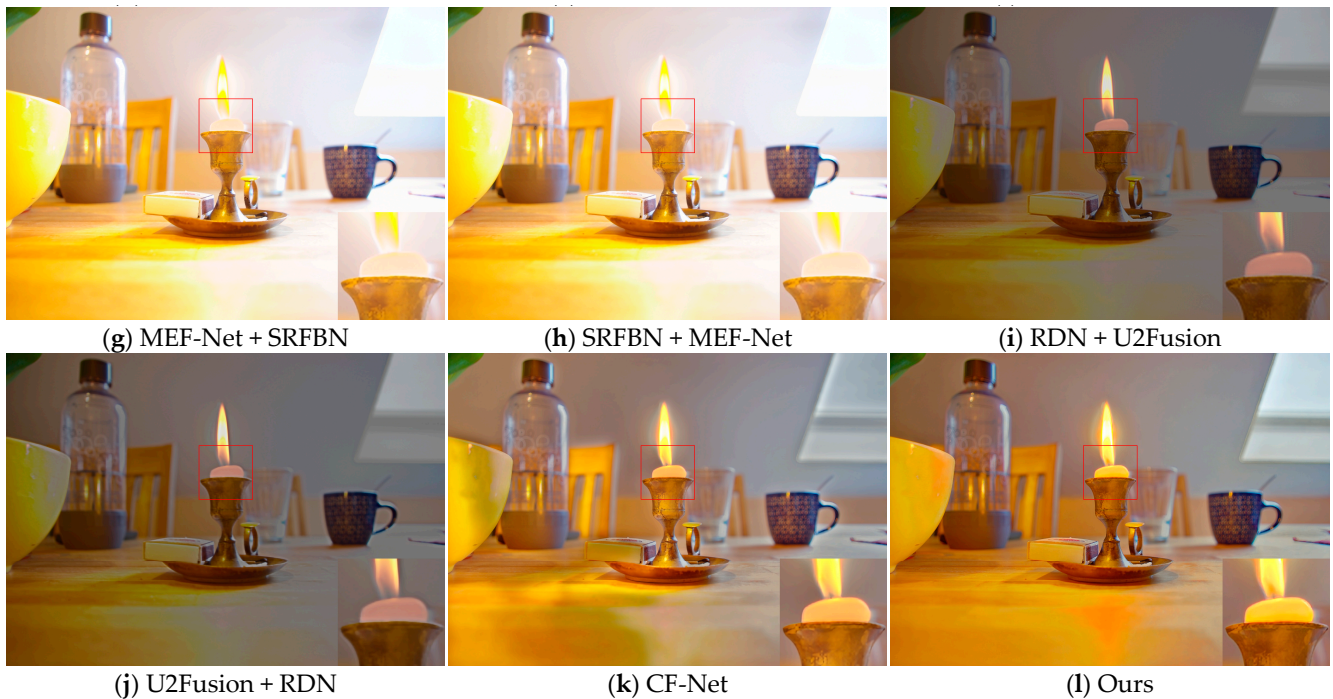


Figure 8. Visual comparison of super-resolution and exposure fusion results from experiments on PQA-MEF candle images with $4\times$ upscaling: (a,b) over-exposed and under-exposed input images, respectively; (c–k) results obtained by the compared methods; and (l) result obtained by our method.

The qualitative findings presented here indicate that our proposed method effectively enhanced the overall color contrast and dynamic range of the produced images. Particularly in the $\times 4$ upscaling results, despite the input low-resolution image having few features, our method enriched the images' textural features and preserved fine details. While our method did not yield the best MEF-SSIM results in the quantitative comparison using the PQA-MEF dataset, the subsequent qualitative comparison demonstrated its superior SR and fusion outcomes.

4.4. Ablation Study

In this subsection, we report a series of experiments conducted to investigate the effects of the proposed attention mechanism, the number of MEFBs, and the output weight on the performance of MEFSR-GAN.

4.4.1. Effect of Attention Mechanism

In an MEFB, the attention mechanism plays a crucial role in enhancing resolution and fusion effects. To demonstrate the efficacy of the proposed attention mechanism, we evaluated MEFSR-GAN's performance on the SICE dataset with and without its attention mechanism. The results are shown in Figure 9, indicating that removing the attention mechanism module led to erroneous halos near features with significant brightness gradient changes. Additionally, as illustrated in Table 5, the inclusion of the attention mechanism resulted in higher PSNR and SSIM values.

Table 5. The effect of the attention mechanism illustrated by evaluating PSNR and SSIM.

| Metric | Without Attention Mechanism | Added Attention Mechanism |
|--------|-----------------------------|---------------------------|
| PSNR | 24.746 | 24.821 |
| SSIM | 0.881 | 0.896 |

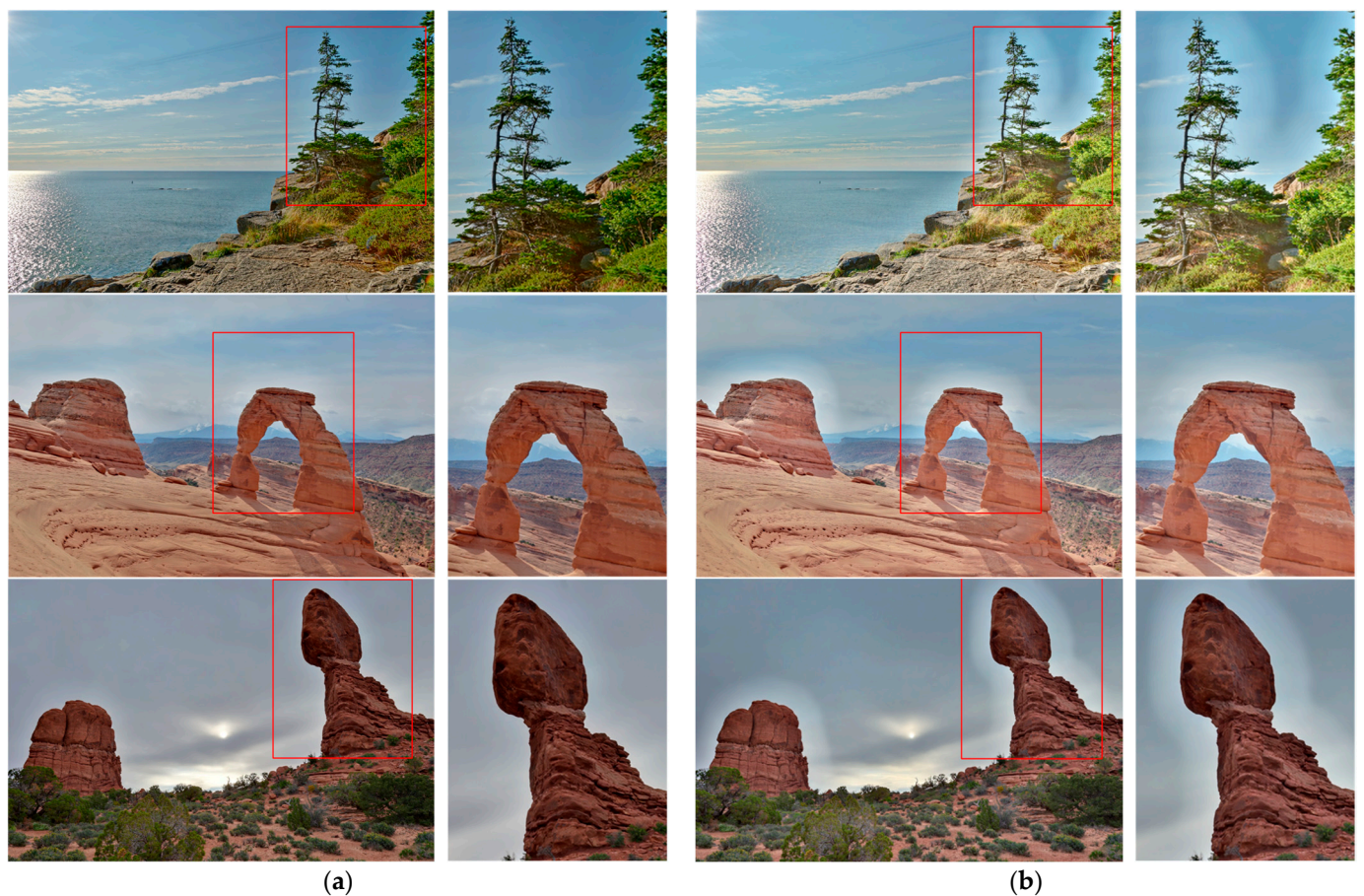


Figure 9. Comparison of visual results obtained when adding or removing the attention mechanism module: (a) with added attention mechanism; (b) without attention mechanism.

4.4.2. Effect of the Number of MEFBs

The results of our ablation experiments, as presented in Table 6, indicated that varying the number of MEFBs had a notable impact on our network's performance. Specifically, we observed that the network achieved optimal PSNR and SSIM scores when utilizing three MEFBs.

Table 6. Effects of different numbers of MEFBs.

| Metric | Num = 2 | Num = 3 | Num = 4 |
|--------|---------|---------|---------|
| PSNR | 23.503 | 24.821 | 22.689 |
| SSIM | 0.885 | 0.896 | 0.876 |

4.4.3. Effect of Output Weight

As demonstrated in Equation (8), the final image quality is influenced by the weighting of the last MEFB. Therefore, the values of the two output weights play a crucial role in determining the overall image quality. To explore their effect, we tested our model with η_o and η_u varying from 0.3 to 0.7. The specific results are presented in Table 7. Based on analysis of the final output images' PSNR and SSIM values, we found that the model achieved the best results when η_o and η_u were both set to 0.5.

Table 7. Effect of output weight value on PSNR and SSIM.

| η_o | η_u | PSNR | SSIM |
|----------|----------|--------|-------|
| 0.3 | 0.7 | 24.801 | 0.895 |
| 0.4 | 0.6 | 24.817 | 0.895 |
| 0.5 | 0.5 | 24.821 | 0.896 |
| 0.6 | 0.4 | 24.813 | 0.896 |
| 0.7 | 0.3 | 24.793 | 0.895 |

5. Discussion

In this section, we provide a comprehensive analysis of prior research and the experimental findings, with the objective of underscoring the significance of the results.

With the SICE dataset, our method achieved a PSNR of 24.821, SSIM of 0.896, and MEF-SSIM of 0.855 in the $2\times$ experiment. For the results of the $4\times$ experiment, our method yielded a PSNR of 21.928, an SSIM of 0.729, and a MEF-SSIM of 0.743. Both sets of results represent the best performances recorded. Similarly, the visualization results further confirm the excellent performance of our network. In Figure 5, the images labeled (c,d) exhibit erroneous black shadows in the sky, while the background noise in images (e,f) is excessive, resulting in unclear details. Additionally, the sky in images (g,h) is over-exposed, and images (i,j) are generally dark, obscuring the dark features on the ground. The peak in image (k) displays an incorrect halo. In contrast, only our method achieved a higher resolution and dynamic range without artifacts or false halos. Figure 7 presents the results for the dynamic scene $4\times$ experiment. Similar to the $2\times$ results, images obtained with methods (c,d) exhibit black artifacts in the sky, along with blurred wave features. Those acquired with methods (e,f) include the presence of uneven color patches and noise. In images (g,h), the sky is over-exposed, while the overall color in images (i,j) appears dim, failing to achieve the desired HDR effect. In image (k), a false halo can be observed at the junction of the sea and sky, contributing to the blurring of wave characteristics. In contrast, our method successfully restored the details of wave characteristics and achieved a higher dynamic range.

In the PQA-MEF dataset experiment, the absence of a ground truth (GT) image precluded a direct comparison of PSNR and SSIM results. Instead, we evaluated the experimental outcomes using MEF-SSIM. Our method achieved an MEF-SSIM score of 0.843 at both $2\times$ and $4\times$ magnification. The resulting value of 0.748 indicates that we did not attain the optimal MEF-SSIM. This limitation arose from the lack of a GT image, as we calculated MEF-SSIM by combining the input low-resolution over-exposed and under-exposed images. Following the image interpolation method, this combined image was used as a surrogate GT for the MEF-SSIM calculation. However, due to the absence of genuine HDR information in both the under-exposed and over-exposed images, the evaluation of MEF results lacked precision. It is important to note that MEF-SSIM may not provide an accurate assessment of an image's color reproduction quality. Consequently, an image that exhibits poor structural detail or inaccurate color reproduction may still achieve a high MEF-SSIM score. However, such images are likely to be perceived as low-quality by the human eye. Specific visualization results can be found in Figures 6 and 8. In Figure 6, black artifacts can be observed above the sun in the images obtained with methods (c,d). Additionally, the background of the hot air balloon letters has not been accurately restored. The solid color sections of the hot air balloons in (e,f) exhibit increased noise, and the letters display incorrect halos. Furthermore, erroneous halos and artifacts are present around the sun in figures (g,h). The overall images in figures (i,j) appear darker, and the hot air balloon letters in figure (k) contain incorrect features. In contrast, our method successfully restored both the letter features and background details of the hot air balloon, while ensuring that the sky remained free of artifacts and false halos.

6. Conclusions

This paper introduces a novel multi-task network that combines multi-exposure fusion and super-resolution reconstruction in a unified framework. The proposed MEFSR-GAN architecture consists of a generator and two discriminators, thus enabling end-to-end processing. The generator comprises under- and over-exposure sub-networks, each incorporating a feature extraction block (FEB), a super-resolution block (SRB), and multiple-exposure feedback blocks (MEFBs). By taking low-resolution under-exposed and over-exposed images as input, the generator extracts features through the FEB, generates high-level features via the SRB, and further refines them through the MEFBs to produce two high-resolution HDR images. The inclusion of a channel attention mechanism in each MEFB enhances image feature details and mitigates halo effects from over-exposure. Since the generated images are evaluated by two discriminators, the generator is encouraged to generate more realistic high-resolution HDR images through simultaneous optimization. Experimental results demonstrate the superiority of our approach over existing methods in terms of both super-resolution accuracy and fusion performance.

Author Contributions: Conceptualization, S.Y.; methodology, S.Y., C.T., G.Z., W.Y. and X.W.; software, S.Y.; validation, S.Y. and K.W.; formal analysis, S.Y.; investigation, C.T.; resources, C.T.; data curation, S.Y.; writing—original draft preparation, S.Y.; writing—review and editing, S.Y., C.T., G.Z., X.W. and K.W.; visualization, S.Y., C.T., W.Y. and K.W.; supervision, C.T. and X.W.; project administration, C.T.; funding acquisition, C.T. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (Grant No. 62205331) and in part by the National Key R&D Plan of China (Grant No. 2022YFF0708500).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lei, F.; Crow, W.T.; Shen, H.; Su, C.-H.; Holmes, T.R.; Parinussa, R.M.; Wang, G. Assessment of the impact of spatial heterogeneity on microwave satellite soil moisture periodic error. *Remote Sens. Environ.* **2018**, *205*, 85–99. [[CrossRef](#)] [[PubMed](#)]
2. Lee, S.-H.; Kim, T.-E.; Choi, J.-S. Correction of radial distortion using a planar checkerboard pattern and its image. *IEEE Trans. Consum. Electron.* **2009**, *55*, 27–33. [[CrossRef](#)]
3. Hsu, W.-Y.; Jian, P.-W. Detail-Enhanced Wavelet Residual Network for Single Image Super-Resolution. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 5016913. [[CrossRef](#)]
4. Wu, W.; Yang, X.; Liu, K.; Liu, Y.; Yan, B.; Hua, H. A new framework for remote sensing image super-resolution: Sparse representation-based method by processing dictionaries with multi-type features. *J. Syst. Arch.* **2016**, *64*, 63–75. [[CrossRef](#)]
5. Ignatov, A.; Kobyshev, N.; Timofte, R.; Vanhoey, K.; Van Gool, L. DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 3297–3305.
6. Timofte, R.; De Smet, V.; Van Gool, L. A plus: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution. In Proceedings of the 12th Asian Conference on Computer Vision (ACCV), Singapore, 1–5 November 2014; Volume 9006, pp. 111–126.
7. Zhou, Y.; Deng, W.; Tong, T.; Gao, Q. Guided Frequency Separation Network for Real-World Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 1722–1731.
8. Ji, X.; Cao, Y.; Tai, Y.; Wang, C.; Li, J.; Huang, F. Real-World Super-Resolution via Kernel Estimation and Noise Injection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 1914–1923.
9. SMAeda, S. Unpaired Image Super-Resolution using Pseudo-Supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 288–297.
10. Yuan, Y.; Liu, S.; Zhang, J.; Zhang, Y.; Dong, C.; Lin, L. Unsupervised Image Super-Resolution Using Cycle-in-Cycle Generative Adversarial Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 814–823. [[CrossRef](#)]
11. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [[CrossRef](#)] [[PubMed](#)]

12. Dong, C.; Loy, C.C.; Tang, X. Accelerating the Super-Resolution Convolutional Neural Network. In *Computer Vision-Eccv 2016*; Li, P., Leibe, B., Eds.; Springer International Publishing: Amsterdam, The Netherlands, 2016; Volume 9906, pp. 391–407.
13. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.P.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
14. Jinno, T.; Okuda, M. Multiple Exposure Fusion for High Dynamic Range Image Acquisition. *IEEE Trans. Image Process.* **2012**, *21*, 358–365. [[CrossRef](#)]
15. Jia, W.; Song, Z.; Li, Z. Multi-Scale Exposure Fusion via Content Adaptive Edge-Preserving Smoothing Pyramids. *IEEE Trans. Consum. Electron.* **2022**, *68*, 317–326. [[CrossRef](#)]
16. Lefevre, S.; Tuia, D.; Wegner, J.D.; Produit, T.; Nassaar, A.S. Toward Seamless Multiview Scene Analysis from Satellite to Street Level. *Proc. IEEE* **2017**, *105*, 1884–1899. [[CrossRef](#)]
17. Yan, Q.; Sun, J.; Li, H.; Zhu, Y.; Zhang, Y. High dynamic range imaging by sparse representation. *Neurocomputing* **2017**, *269*, 160–169. [[CrossRef](#)]
18. Yang, Z.; Chen, Y.; Le, Z.; Ma, Y. GANFuse: A novel multi-exposure image fusion method based on generative adversarial networks. *Neural Comput. Appl.* **2021**, *33*, 6133–6145. [[CrossRef](#)]
19. Abed, F.; Khan, I.R.; Rahardja, S. A New Four-Channel Format for Encoding of HDR Images. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2018**, *E101A*, 512–515. [[CrossRef](#)]
20. Shermeyer, J.; Van Etten, A. The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Long Beach, CA, USA, 15–20 June 2019; pp. 1432–1441.
21. Deng, X.; Zhang, Y.; Xu, M.; Gu, S.; Duan, Y. Deep Coupled Feedback Network for Joint Exposure Fusion and Image Super-Resolution. *IEEE Trans. Image Process.* **2021**, *30*, 3098–3112. [[CrossRef](#)] [[PubMed](#)]
22. Hassan, M.; Wang, Y.; Pang, W.; Wang, D.; Li, D.; Zhou, Y.; Xu, D. IPAS-Net: A deep-learning model for generating high-fidelity shoeprints from low-quality images with no natural references. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 2743–2757. [[CrossRef](#)]
23. He, Z.; Jin, Z.; Zhao, Y. SRDRL: A Blind Super-Resolution Framework With Degradation Reconstruction Loss. *IEEE Trans. Multimed.* **2021**, *24*, 2877–2889. [[CrossRef](#)]
24. Li, Y.; Wang, Y.; Li, Y.; Jiao, L.; Zhang, X.; Stolkin, R. Single image super-resolution reconstruction based on genetic algorithm and regularization prior model. *Inf. Sci.* **2016**, *372*, 196–207. [[CrossRef](#)]
25. Irani, M.; Peleg, S. Improving resolution by image registration. *CVGIP Graph. Models Image Process.* **1991**, *53*, 231–239. [[CrossRef](#)]
26. Ur, H.; Gross, D. Improved resolution from subpixel shifted pictures. *CVGIP Graph. Model. Image Process.* **1992**, *54*, 181–186. [[CrossRef](#)]
27. Schultz, R.; Stevenson, R. A Bayesian approach to image expansion for improved definition. *IEEE Trans. Image Process.* **1994**, *3*, 233–242. [[CrossRef](#)]
28. Schultz, R.; Stevenson, R. Extraction of high-resolution frames from video sequences. *IEEE Trans. Image Process.* **1996**, *5*, 996–1011. [[CrossRef](#)]
29. Elad, M.; Feuer, A. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. Image Process.* **1997**, *6*, 1646–1658. [[CrossRef](#)]
30. Lertrattanapanich, S.; Bose, N. High resolution image formation from low resolution frames using delaunay triangulation. *IEEE Trans. Image Process.* **2002**, *11*, 1427–1441. [[CrossRef](#)] [[PubMed](#)]
31. Freeman, W.; Jones, T.; Pasztor, E. Example-based super-resolution. *IEEE Comput. Graph. Appl.* **2002**, *22*, 56–65. [[CrossRef](#)]
32. Yang, J.; Wright, J.; Huang, T.; Ma, Y. Image super-resolution as sparse representation of raw image patches. In *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
33. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image Super-Resolution Via Sparse Representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873. [[CrossRef](#)]
34. Timofte, R.; De Smet, V.; Van Gool, L. Anchored Neighborhood Regression for Fast Example-Based Super-Resolution. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Sydney, Australia, 1–8 December 2013; pp. 1920–1927.
35. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a Deep Convolutional Network for Image Super-Resolution. In *Proceedings of the ECCV 2014*, Zurich, Switzerland, 6–12 September 2014; pp. 184–199. [[CrossRef](#)]
36. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
37. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 1637–1645.
38. Tai, Y.; Yang, J.; Liu, X.; Xu, C. MemNet: A Persistent Memory Network for Image Restoration. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 22–29 October 2017; pp. 4549–4557.
39. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.

40. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; Volume 11211, pp. 294–310.
41. Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; Dong, C. Activating More Pixels in Image Super-Resolution Transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 22367–22377.
42. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. SwinIR: Image Restoration Using Swin Transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Montreal, BC, Canada, 11–17 October 2021; pp. 1833–1844.
43. Chen, Y.; Liu, S.; Wang, X. Learning continuous image representation with local implicit image function. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 8628–8638.
44. Burt, P.J.; Adelson, E.H. The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.* **1983**, *31*, 532–540. [[CrossRef](#)]
45. Mertens, T.; Kautz, J.; Van Reeth, F. Exposure fusion. In Proceedings of the 15th Pacific Conference on Computer Graphics and Applications (PG'07), Maui, HI, USA, 29 October–2 November 2007; pp. 382–390.
46. Goshtasby, A.A.; Nikolov, S. Image fusion: Advances in the state of the art. *Inf. Fusion* **2007**, *8*, 114–118. [[CrossRef](#)]
47. Burt, P.J.; Kolczynski, R.J. Enhanced image capture through fusion. In Proceedings of the 1993 (4th) International Conference on Computer Vision, Berlin, Germany, 11–14 May 1993; pp. 173–182.
48. Goshtasby, A.A. Fusion of multi-exposure images. *Image Vis. Comput.* **2005**, *23*, 611–618. [[CrossRef](#)]
49. Ma, K.; Wang, Z. Multi-exposure image fusion: A patch-wise approach. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 1717–1721.
50. Liu, Y.; Liu, S.; Wang, Z. Multi-focus image fusion with dense SIFT. *Inf. Fusion* **2015**, *23*, 139–155. [[CrossRef](#)]
51. Lee, S.H.; Park, J.S.; Cho, N.I. A multi-exposure image fusion based on the adaptive weights reflecting the relative pixel intensity and global gradient. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 1737–1741.
52. Prabhakar, K.R.; Srikar, V.S.; Babu, R.V. DeepFuse: A Deep Unsupervised Approach for Exposure Fusion with Extreme Exposure Image Pairs. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4724–4732. [[CrossRef](#)]
53. Zhang, Y.; Liu, Y.; Sun, P.; Yan, H.; Zhao, X.; Zhang, L. IFCNN: A general image fusion framework based on convolutional neural network. *Inf. Fusion* **2020**, *54*, 99–118. [[CrossRef](#)]
54. Li, H.; Zhang, L. Multi-exposure fusion with CNN features. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 1723–1727.
55. Xu, H.; Ma, J.; Zhang, X.-P. MEF-GAN: Multi-Exposure Image Fusion via Generative Adversarial Networks. *IEEE Trans. Image Process.* **2020**, *29*, 7203–7216. [[CrossRef](#)]
56. Xu, H.; Ma, J.; Le, Z.; Jiang, J.; Guo, X. FusionDn: A unified densely connected network for image fusion. In Proceedings of the 34th AAAI Conference on Artificial Intelligence/32nd Innovative Applications of Artificial Intelligence Conference/10th AAAI Symposium on Educational Advances in Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12484–12491.
57. Xu, H.; Ma, J.; Jiang, J.; Guo, X.; Ling, H. U2Fusion: A Unified Unsupervised Image Fusion Network. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 502–518. [[CrossRef](#)]
58. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C.C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In Proceedings of the 15th European Conference on Computer Vision, ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 63–79.
59. Haris, M.; Shakhnarovich, G.; Ukita, N. Deep Back-Projection Networks For Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 1664–1673.
60. Li, Z.; Yang, J.; Liu, Z.; Yang, X.; Jeon, G.; Wu, W. Feedback Network for Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 3862–3871.
61. Jolicœur-Martineau, A. The relativistic discriminator: A key element missing from standard GAN. *arXiv* **2018**, arXiv:1807.00734.
62. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
63. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
64. Cai, J.; Gu, S.; Zhang, L. Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. *IEEE Trans. Image Process.* **2018**, *27*, 2049–2062. [[CrossRef](#)]
65. Ma, K.; Zeng, K.; Wang, Z. Perceptual Quality Assessment for Multi-Exposure Image Fusion. *IEEE Trans. Image Process.* **2015**, *24*, 3345–3356. [[CrossRef](#)]
66. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual Dense Network for Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.
67. Ma, K.; Duanmu, Z.; Zhu, H.; Fang, Y.; Wang, Z. Deep Guided Learning for Fast Multi-Exposure Image Fusion. *IEEE Trans. Image Process.* **2020**, *29*, 2808–2819. [[CrossRef](#)]

-
68. Li, H.; Ma, K.; Yong, H.; Zhang, L. Fast Multi-Scale Structural Patch Decomposition for Multi-Exposure Image Fusion. *IEEE Trans. Image Process.* **2020**, *29*, 5805–5816. [[CrossRef](#)]
 69. Ma, K.; Duanmu, Z.; Yeganeh, H.; Wang, Z. Multi-exposure image fusion by optimizing a structural similarity index. *IEEE Trans. Comput. Imaging* **2017**, *4*, 60–72. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.