remote sensing

MDPI

Article

ERS-HDRI: Event-Based Remote Sensing HDR Imaging

Xiaopeng Li, Shuaibo Cheng, Zhaoyuan Zeng, Chen Zhao and Cien Fan *

The School of Electronic Information, Wuhan University, Wuhan 430072, China; xiaopengli2014@whu.edu.cn (X.L.); cheng2018@whu.edu.cn (S.C.); zhaoyuan.zeng@whu.edu.cn (Z.Z.); zhaoc@whu.edu.cn (C.Z.)
* Correspondence: fce@whu.edu.cn

Abstract: High dynamic range imaging (HDRI) is an essential task in remote sensing, enhancing low dynamic range (LDR) remote sensing images and benefiting downstream tasks, such as object detection and image segmentation. However, conventional frame-based HDRI methods may encounter challenges in real-world scenarios due to the limited information inherent in a single image captured by conventional cameras. In this paper, an event-based remote sensing HDR imaging framework is proposed to address this problem, denoted as ERS-HDRI, which reconstructs the remote sensing HDR image from a single-exposure LDR image and its concurrent event streams. The proposed ERS-HDRI leverages a coarse-to-fine framework, incorporating the event-based dynamic range enhancement (E-DRE) network and the gradient-enhanced HDR reconstruction (G-HDRR) network. Specifically, to efficiently achieve dynamic range fusion from different domains, the E-DRE network is designed to extract the dynamic range features from LDR frames and events and perform intra- and cross-attention operations to adaptively fuse multi-modal data. A denoise network and a dense feature fusion network are then employed for the generation of the coarse, clean HDR image. Then, the G-HDRR network, with its gradient enhancement module and multiscale fusion module, performs structure enforcement on the coarse HDR image and generates a fine informative HDR image. In addition, this work introduces a specialized hybrid imaging system and a novel, real-world event-based remote sensing HDRI dataset that contains aligned remote sensing LDR images, remote sensing HDR images, and concurrent event streams for evaluation. Comprehensive experiments have demonstrated the effectiveness of the proposed method. Specifically, it improves state-of-the-art PSNR by about 30% and the SSIM score by about 9% on the real-world dataset.

Keywords: HDR imaging; multi-modal fusion; remote sensing image processing; event camera; machine learning

Citation: Li, X.; Cheng, S.; Zeng, Z.; Zhao, C.; Fan, C. ERS-HDRI: Event-Based Remote Sensing HDR Imaging. *Remote Sens.* **2024**, *16*, 437. https://doi.org/10.3390/rs16030437

Academic Editor: Andrea Garzelli

Received: 19 December 2023
Revised: 19 January 2024
Accepted: 20 January 2024
Published: 23 January 2024

1. Introduction

Remote sensing photography through airplanes is important for earth observation, aiming to record the spatial information of large areas on Earth [1–4]. However, due to the unexpected light conditions, remote sensing images captured in the real world always suffer from low dynamic range, leading to incomplete scene information [5]. Existing approaches handle high dynamic range (HDR) reconstruction by leveraging multiple low dynamic range (LDR) images with different exposures, i.e., multi-exposure high dynamic range imaging (HDRI) [6,7], or a single LDR image, i.e., single-exposure HDRI [8–11]. Nevertheless, multi-exposure HDRI often suffers from ghosting caused by moving objects [10,12,13]. Even though single-exposure HDRI is more efficient and immune to ghosting effects, its limited information presents a challenge for HDRI [8,14,15], rendering the problem ill-posed, as shown in Figure 1a.

In recent years, event cameras have shown great advantages in HDR imaging [16]. Since the photoreceptors of pixels measure intensity changes asynchronously in the logarithmic domain [16,17], event cameras achieve a high dynamic range (>120 dB) and can compensate for saturated regions in LDR images captured by conventional cameras.

However, the recovery of high-quality remote sensing HDR images with the aid of event streams is still a challenging problem due to the following issues:

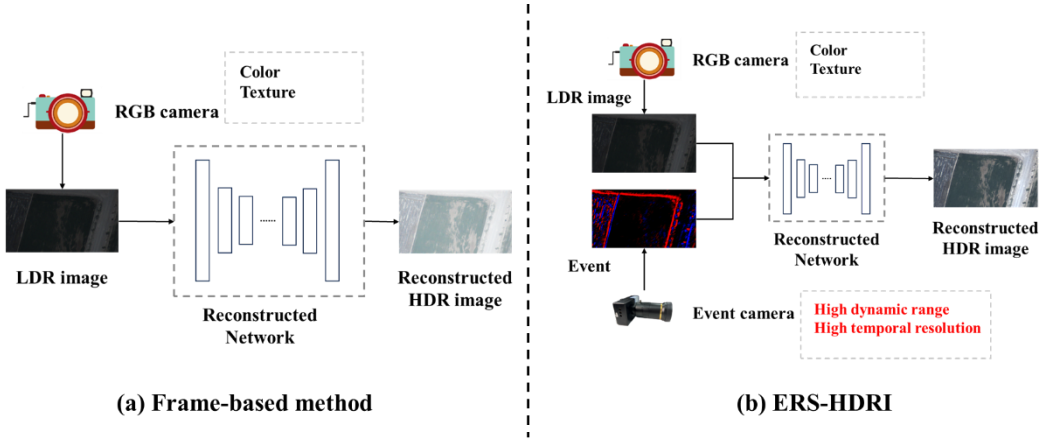

Figure 1. Different HDR reconstruction methods. (**a**): previous frame-based method. (**b**): our proposed ERS-HDRI.

- **Domain Gap.** Conventional RGB cameras continuously capture frames by integrating brightness and then generating color frames. In contrast, event cameras operate on a completely different principle, detecting and transmitting changes in luminance, resulting in an asynchronous stream of events [16]. The substantial distinction in imaging mechanisms between conventional cameras and event cameras gives rise to a considerable domain gap between optical images and event streams, preventing their efficient integration. Existing event-guided HDR imaging methods have successfully integrated image frames and event streams by introducing exposure mask attention [18,19]. However, the exposure mask is generated through threshold segmentation and cannot be learned according to different environments, resulting in an inability to perfectly adapt to diverse scenes. Therefore, how to narrow the domain gap and implement adaptive fusion between optical images and event streams is still an open problem in event-guided HDRI tasks.
- **Light attenuation.** The structures within low dynamic range (LDR) frames typically exhibit weakening in under-/over-exposed regions. Even though event cameras are able to sense structure information at contrast edges, their effectiveness in capturing detailed information diminishes when operating at high altitudes. This limitation arises due to the decrease in light intensity with increasing distance; as a result, the event camera's perception of brightness changes often fails to reach the event triggering threshold when capturing images at high altitudes [16], making it difficult for the event camera to capture complex details. Therefore, it is challenging to reconstruct informative structures in badly exposed remote sensing images with events captured at high altitudes.

Previous event-based HDRI methods have attempted to leverage event streams to guide LDR image enhancement by designing two-stage or end-to-end networks in ground photography [18,20]. However, when it comes to remote sensing HDRI, the increased sparsity of events aggravates the challenge of bridging the domain gap between LDR frames and events for fusion. Moreover, the impact of light attenuation hinders the effectiveness of these methods, resulting in the loss of details and distortion in the reconstructed images.

In this paper, an event-based remote sensing HDR imaging framework, i.e., ERS-HDRI, is proposed to address these problems; it takes a remote sensing LDR image and its concurrent events as input, employing a coarse-to-fine strategy to reconstruct an informative HDR image. ERS-HDRI incorporates two networks: the event-based dynamic range enhancement (E-DRE) network and the gradient-enhanced HDR reconstruction (G-HDRR) network; they perform coarse and fine HDR reconstruction, respectively. Particularly, the

E-DRE network reconstructs the missing information in under-/over-exposed regions by integrating events, where the LDR image and events undergo adaptive fusion through the introduction of the intra- and cross-attention (ICA) module. To handle the noise in under-exposed regions, a multiscale denoising module is introduced into the E-DRE network, followed by the dense feature fusion (DFF) module to enhance reconstruction. In the G-HDRR network, the gradient enhancement (GE) module is leveraged to excavate the structure map from image gradients and event frames, which can be used to enhance textures in low-contrast regions. Additionally, a novel remote sensing event-based HDRI dataset composed of both synthetic data and real-world data is conducted to evaluate our proposed method. Comprehensive experiments demonstrate the superior performance of our approach over state-of-the-art methodologies.

To summarize, the contributions of this work are as follows:

- We introduce an event-based HDRI framework for remote sensing HDR image reconstruction, which integrates LDR frames with event streams.
- We implement a coarse-to-fine strategy that efficiently achieves dynamic range enhancement and structure enhancement, where both the domain gap problem and the light attenuation problem are alleviated.
- We present a hybrid imaging system with a conventional optical camera and an event camera; moreover, we present a novel remote sensing event-based HDRI dataset that contains aligned LDR images, HDR images, and concurrent event streams.

The remainder of the paper is organized as follows. Section 2 reviews the related works including the single-exposure HDR reconstruction, event-based HDR reconstruction, and remote sensing image enhancement. Section 3 presents the problem formulation of the event-based remote sensing HDRI task and describes the details of the proposed ERS-HDRI, including the network architecture and optimization strategy. Section 4 details the hybrid imaging system and the event-based remote sensing HDRI dataset. Finally, this work evaluates the performance of ERS-HDRI on the proposed dataset in Section 5 and concludes in Section 6. Limitations and future work are presented in Section 7.

2. Related Work

2.1. Framed-Based HDR Reconstruction

The framed-based HDR reconstruction is mainly composed of two methodologies, i.e., multi-image HDR reconstruction and single-image HDR reconstruction. The former generates the HDR images through the fusion of a stack of LDR images, each captured at distinct exposure times of the same scene. Although certain efforts in this domain have yielded commendable results, limitations, e.g., significant delays, arise due to dependencies on specific software/hardware technologies, which affect the timeliness of remote sensing imaging in specific scenarios. Therefore, we focus on single-exposure HDR reconstruction methods, which aim to reconstruct missing details within saturated regions with a single LDR input.

Traditional approaches estimate the density of light sources to expand the dynamic range or conduct the cross-bilateral filter to enhance the input LDR images [21–24]. However, processing the diverse and complex semantic information inherent in various scenarios proves to be a huge challenge.

Recently, with the release of several datasets, CNN-based methods have shown great performance. Eilertsen et al. [25] presented HDRCNN to recover missing details in the over-exposed regions. Marnerides et al. designed a multiscale autoencoder architecture, i.e., ExpandNet, aiming to learn different levels of details from an LDR image. However, it is noteworthy that these approaches have tended to ignore the presence of quantization artifacts and noise. Santos et al. [26] contributed to the field by introducing masked features and perceptual loss, which effectively mitigate ambiguity and halo artifacts in HDR images. Nevertheless, this may result in the reconstruction of saturated areas with inaccuracies in color representations in some cases. On the contrary, Liu et al. attempted to learn LDR-to-HDR mapping by reversing the camera pipeline, including dequantization,

linearization, and hallucination [27], demonstrating remarkable performance. Chen et al. introduced a spatially dynamic encoder–decoder network [11], facilitating a robust learning framework for HDR reconstruction that incorporates denoising and dequantization. Akhil et al. enhanced the network's representation capacity through the adoption of distinct dense connection structures [28]. Moreover, Sharif et al. [29] proposed a two-stage learning-based method without hardware information, such as the camera response function (CRF) and exposure settings. Instead of employing multiple networks, Wang et al. presented a unified network based on the imaging process, by integrating LDR-to-HDR imaging knowledge into an UNet architecture [30]. Despite the considerable advances, single-image HDR reconstruction remains a complex and ill-posed problem, primarily due to the challenge of addressing missing information in under-/over-exposed regions with limited information.

On the other hand, inspired by multi-exposure approaches, certain works have also attempted to reconstruct HDR from a single image via the prediction of multiple exposure images [31]. This approach provides more fine-grained control over details and ensures robust HDR recovery under various lighting conditions. However, it is important to note that generating images with distinct exposures is difficult. Moreover, all the information still comes from a single LDR image, leading to limited useful information.

2.2. Event-Based HDR Reconstruction

Event cameras represent a groundbreaking class of bio-inspired neuromorphic sensors, presenting a paradigm shift in the acquisition of visual information. Diverging from the conventional approach of measuring the intensity of each pixel, event cameras asynchronously detect changes in scene radiance, providing remarkable structural texture. The dynamic range of traditional frame-based cameras is limited as they need to encode scene intensities into a fixed number of bits, while event cameras do not encode absolute intensity levels, and will not saturate in extreme lighting conditions. Consequently, event cameras have a substantially higher dynamic range (140 dB vs. 60 dB) [18], rendering them promising and advantageous for HDR imaging.

Given the capacity of event cameras to capture additional scene details in areas poorly exposed in LDR scenarios, numerous works have attempted to reconstruct intensity images solely from event data. Belbachir et al. first handled this challenge within the context of known camera motions [32]. Bardow et al. expanded this to general cameras by estimating joint intensity images and optical flow [33]. By leveraging a recurrent neural network (RNN) for video reconstruction, Rebecq et al. proposed EVDI and obtained promising results [34]. Recently, a series of CNN-based networks have emerged [35–37]. Liang et al. [38] innovatively incorporated the diffusion model into the reconstruction pipeline to remove artifacts and blur in reconstructed images, achieving high-quality results. However, the results of these works are typically low resolution and grayscale, constraining their applicability in diverse scenarios. Furthermore, due to the lack of absolute intensity information and varying contrast thresholds, these approaches frequently fail to provide highly detailed reconstructions.

To leverage the full spectrum of information in the LDR and the event data, some approaches have employed various strategies to utilize events in guiding the LDR-to-HDR mapping. Han et al. [18] proposed a multi-modal camera system and learning framework for HDR, incorporating LDR and the intensity map generated by E2VID [34] as inputs. However, their approach, which does not allow for end-to-end model optimization, resulted in sub-optimal solutions. Messikommer et al. [20] first combined bracketed LDR images and synchronized events for HDR imaging, demonstrating enhanced robustness in handling noise and ghosts. Richard et al. [39] introduced a novel event-to-image feature distillation module, directly transforming event features into the image feature space without relying on an intermediary intensity image. Yang et al. [40] presented a multi-modal learning framework for reconstructing HDR videos from hybrid inputs of LDR videos and events. However, these approaches face challenges in achieving satisfactory generalization results

when applied to remote sensing datasets. We attribute this limitation to the specificity of the data and network structure.

2.3. Remote Sensing Image Enhancement

Remote sensing images have been widely applied in object detection [1,41], image semantic segmentation [42], image classification [43], and change detection [44]. However, the captured remote sensing images under diverse light conditions, such as overexposure and underexposure, often suffer from low dynamics and noise [45,46]. This, in turn, hampers the extraction of information for subsequent tasks.

Recent works have attempted to address this problem by performing remote sensing HDR reconstruction and denoising, respectively. For remote sensing HDR imaging, aimed at HDR reconstruction of low-illumination remote sensing images, Zhang et al. [5] introduced a dynamic long–short-range structure. The network augments the lighting enhancement module, capturing both long-distance and short-distance structural information in LI images, thereby improving the generalization capacity of the method. However, it only focuses on processing under-exposed images, while reconstruction from over-exposed remote sensing images has been less studied. The research on remote sensing image denoising is relatively extensive. Han et al. [47] presented a novel remote sensing image denoising network (RSIDNet), comprising a multiscale feature extraction module, a global feature fusion block, and a noisy image reconstruction block. Synergistic integration of these modules significantly enhances the capabilities of feature extraction, preserving detailed information more effectively. Huang et al. [48] proposed a deep gradient descent network, which can recognize structures from images degraded by additive white Gaussian noise, producing competitive denoising performance. Several studies have integrated image enhancement with additional tasks within the remote sensing domain. Wang et al. [49] introduced a novel cross-modal interactive fusion method to enhance the interactivity of modal fusion by combining multisource information, greatly improving classification accuracy. Xi et al. [50] incorporated an anti-label-noise network framework into semantic segmentation, enhancing the model's robustness by mitigating label noise. However, to the best of our knowledge, scant attention has been directed toward investigating high dynamic range imaging of remote sensing images under arbitrary exposure conditions with noise.

Although one can address the image enhancement task under abnormal lighting by cascading the dynamic range enhancement method and image denoising method mentioned above, sub-optimal problems often arise due to error accumulation. Therefore, this paper considers conducting dynamic range reconstruction and image denoising synchronously.

3. Methods

In this section, the formulation of the event-based remote sensing HDRI problem is presented in Section 3.1. Then, our proposed event-based remote sensing HDR imaging (ERS-HDRI) framework is introduced in Section 3.2, followed by the optimization strategy in Section 3.3.

3.1. Problem Formulation

Given a remote sensing LDR image L defined over $L \triangleq \mathcal{W} + \mathcal{B}$ with $\mathcal{W}$ and $\mathcal{B}$ denoting well-exposed and badly-exposed regions, respectively, frame-based HDR imaging methods recover the HDR image $\hat{I}$ by learning semantically meaningful information of well-exposed regions $L_{\mathcal{W}}$ and the essential distributions of the total image L.

$$\hat{I} = \mathcal{F}(L_{\mathcal{W}}; \mathcal{D}(L)), \tag{1}$$

where $\mathcal{F}$ represents HDR imaging networks and $\mathcal{D}(L)$ denotes the distributions of input LDR images.

However, when L is severely degraded, i.e., with extremely under-/over-exposed regions, the information of $L_{\mathcal{W}}$ is not enough to support the reconstruction of badly exposed

regions, and the weakened distributions $\mathcal{D}(L)$ aggravate the difficulty of this task. Thanks to the high dynamic range of event cameras, the event streams can record more information in cases of inappropriate exposure, thus offering a valuable resource to compensate for the remote sensing HDRI task.

The objective of our event-based remote sensing HDR imaging is to reconstruct the HDR image $\hat{I}$ from a blurry, low dynamic range image L and the concurrent event streams $\mathcal{E}_\mathcal{T} \triangleq \{\mathbf{x}, t, p\}$, which are triggered inside $\mathcal{T}$, where $\mathbf{x}$, t, and p denote the position, timestamp, and polarity of the event. The reconstruction process can be formulated through an operator $\mathcal{G}$ conditioned by both the fused dynamic range and structures, i.e.,

$$\hat{I} = \mathcal{G}(\mathcal{D}(\mathcal{A}(L, \mathcal{E}_\mathcal{T})), \mathcal{S}(L, \mathcal{E}_\mathcal{T})), \tag{2}$$

where $\mathcal{D}$ represents the dynamic range feature fusion operation. Considering the different distributions between L and $\mathcal{E}_\mathcal{T}$, the feature adaptive fusion $\mathcal{A}$ is introduced into $\mathcal{D}$ to realize the adaptive fusion multi-modal features. $\mathcal{S}$ is a structure enhancement operation that excavates the edge information to maintain the structure in the recovered HDR images.

3.2. Network Architecture

The proposed event-based framework aims to reconstruct the remote sensing HDR image from an LDR image and its concurrent event streams, based on Equation (2), termed ERS-HDRI. The overview pipeline is composed of two sub-networks, i.e., the event-based dynamic range enhancement network (E-DRE) and the gradient-enhanced HDR reconstruction network (G-HDRR), as shown in Figure 2. Based on the coarse-to-fine strategy, E-DRE firstly performs coarse HDR imaging on the LDR image L with the aid of event streams $\mathcal{E}_\mathcal{T}$ and reconstructs the HDR image $\hat{I}_{coarse}$. Secondly, by excavating the structure attention map from the gradient of the coarse HDR image and short temporal event frame, the G-HDRR network enhances the structures in low-contrast regions and reconstructs the final informative and visually pleasing sharp HDR image $\hat{I}_{fine}$.

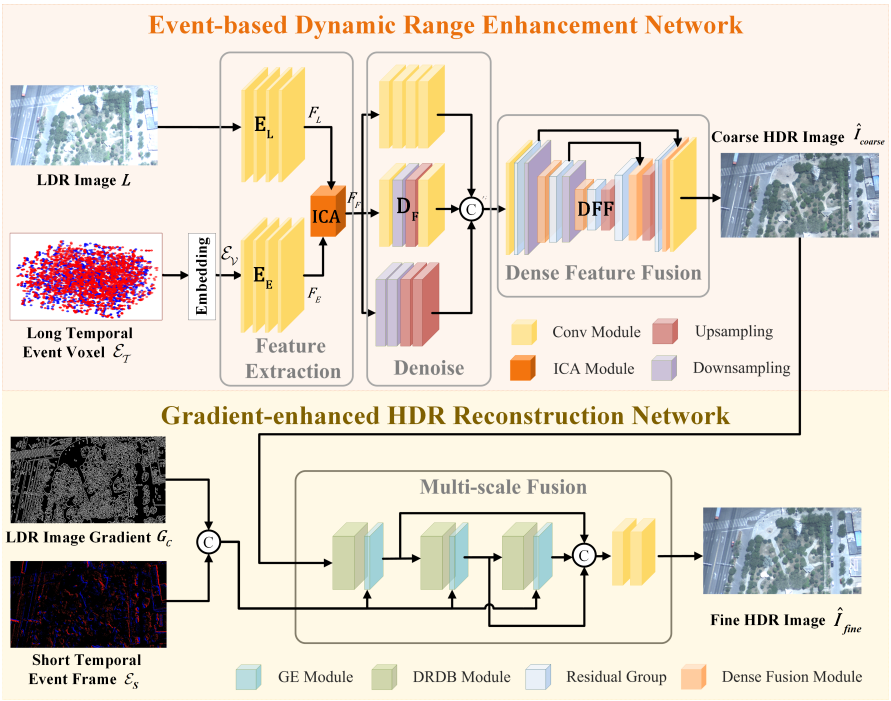

Figure 2. An illustration of our event-based remote sensing HDR imaging framework, i.e., ERS-HDRI, which is composed of the event-based dynamic range enhancement (E-DRE) network and gradient-enhanced HDR reconstruction (G-HDRR) network.

3.2.1. Event-Based Dynamic Range Enhancement Network

The event-based dynamic range enhancement network (E-DRE) aims to learn the HDR imaging function $\mathcal{D}(\mathcal{A}(L, \mathcal{E}_\mathcal{T}))$. Since events are powerful at recording high dynamic range scenes, they can recover the actual textures in the saturated regions of the LDR images. Therefore, a coarse dynamic range fusion network, composed of feature extraction, denoising, and dense feature fusion, is conducted to extract the high dynamic range information from both the LDR image and event streams, and then reconstructs a coarse clean high dynamic range image, as shown in Figure 2.

The purpose of the feature extraction process is to acquire meaningful information from both the LDR image and events, encompassing dynamic range, structures, and color. To enhance the processability of event streams by convolutional neural networks, they are initially embedded into voxel grids. Specifically, when dealing with the event stream $\mathcal{E}_\mathcal{T}$, it is divided into 5 temporal bins, and then the events within each bin are merged into $h \times w$ tensors. Consequently, the event streams are represented by a $5 \times h \times w$ tensor, denoted as $\mathcal{E}_\mathcal{V}$. Considering the notable disparity in distributions between the LDR image L and the event voxel $\mathcal{E}_\mathcal{V}$, direct fusion at the image level becomes challenging due to the domain gap between two types of multi-modal data. To address this, two encoders, i.e., $\mathrm{E_L}$ and $\mathrm{E_E}$, are employed to embed L and $\mathcal{E}_\mathcal{V}$ into the feature space, yielding distinctive features, denoted as F_L and F_E.

$$F_L = \mathrm{E_L}(L),\ F_E = \mathrm{E_E}(\mathcal{E}_\mathcal{V}). \tag{3}$$

Note that the encoders are composed of cascaded convolution blocks, each incorporating a convolution layer, batch normalization layer, and ReLU activation function.

To facilitate adaptive fusion between multi-modal features, our proposed framework incorporates considerations for both the intra-relationship within each modal feature and the cross-relationship between multi-modal features during the fusion process. Thus, the intra- and cross-attention (ICA) module is proposed to merge deep features derived from different modalities, as depicted in Figure 3. Specifically, given the LDR image feature, F_L, and event feature, F_E, two parallel feature attention operations are executed to learn crucial information within each modal branch.

For the intra-attention operation, one convolution block is used to generate the image content feature, $L_{content}$, followed by a dual attention module consisting of channel attention and spatial attention operations, facilitating intra-fusion. The intra-attention operation enables the network to extract valuable information from single-modal data, mitigating interference from degraded information. The process can be expressed as follows:

$$L'_{content} = \text{Dual-Attention}(L_{content}),\ E'_{content} = \text{Dual-Attention}(E_{content}). \tag{4}$$

For the cross-attention operation, the cascaded convolution blocks are employed to generate the image attention feature map M_L and event attention feature map M_E from F_L and F_E, respectively. Then, M_L and M_E are leveraged to refine F_L and F_E through dot multiplication and generate F'_L and F'_E, respectively, aiming to provide weighted adjustment across the modalities. After that, the summation is carried out between the features generated from the intra-attention and cross-attention modules, obtaining the fused features by performing the concatenation operation,

$$F_F = \text{Concat}((F'_L + L'_{content}), (F'_E + E'_{content})). \tag{5}$$

As noise is prevalent in the under-exposed regions of remote sensing LDR images, a denoise module is introduced to address degradation at the feature level. By designing denoising networks $\mathrm{D_F}$ of different scales and subsequently fusing their feature outputs, the impact of random noise on restoration results is reduced in different receptive fields. Subsequently, the dense feature fusion (DFF) network is deployed for coarse HDR image reconstruction, where the dense fusion modules [51] serve as the main components, en-

suring a sufficient connection between non-adjacent levels of features and rectifying the missing spatial information during downsampling and upsampling. The denoising and HDR reconstruction process can be expressed as follows:

$$\hat{I}_{coarse} = \mathrm{DFF}(\mathrm{D_F}(F_F)). \tag{6}$$

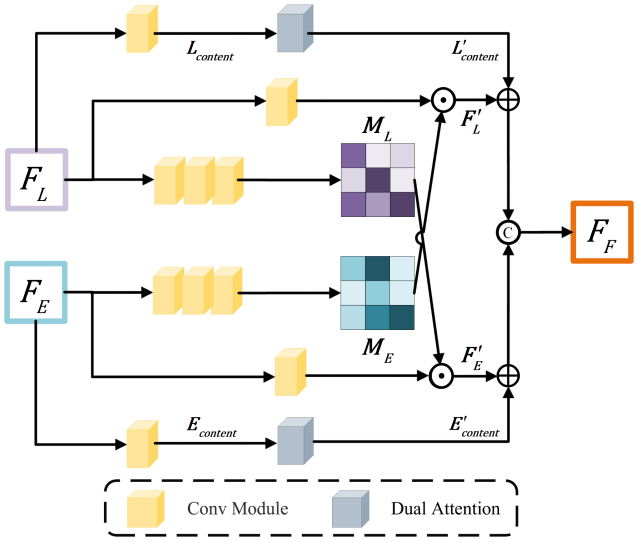

Figure 3. The implementation of our proposed intra- and cross-attention (ICA) module for the LDR image feature and event feature fusion.

3.2.2. Gradient-Enhanced HDR Reconstruction Module

Since remote sensing data are captured at high altitudes, the obtained event streams always lack information at regions with small brightness differences due to light attenuation, which prevents the effective reconstruction of low-contrast regions. To mitigate this problem, the gradient-enhanced HDR reconstruction (G-HDRR) network is introduced to reconstruct a more informative HDR image, $\hat{I}_{fine}$, by leveraging the structure attention map derived from both the gradient, G_C, of the coarse HDR image and the event frame, to reinforce the structures. The process is formulated as follows:

$$\hat{I}_{fine} = \text{G-HDRR}(\hat{I}_{coarse}, G_C, \mathcal{E}_S) \tag{7}$$

where $\mathcal{E}_S$ denotes the short temporal event frame, which is selected from the middle voxel bin of $\mathcal{E}_V$ and contains the information of a short temporal span.

The G-HDRR network is mainly composed of a multiscale fusion module that comprises the dense residual blocks with dilated convolution, i.e., DRDB [13], where multiscale feature extraction and fusion are processed by applying dilated convolution with the varying receptive fields. Effective feature fusion is achieved by incorporating both local residual skip connections within a DRDB block and global residual skip connections between the three DRDB blocks. To enhance the structures during the HDR reconstruction process, the gradient enhancement (GE) module is employed to encode the structure attention map, inspired by AIND [52], which guides denoising with the noise estimation map, as shown in Figure 4. Specifically, the GE module takes the concatenation of the coarse HDR image gradient G_C and short temporal event frame $\mathcal{E}_S$ as input and leverages cascaded learnable convolutional blocks to encode positional information related to the structure. By modulating the input features through scaling and shifting operations, one can enhance the details in the low-contrast regions.

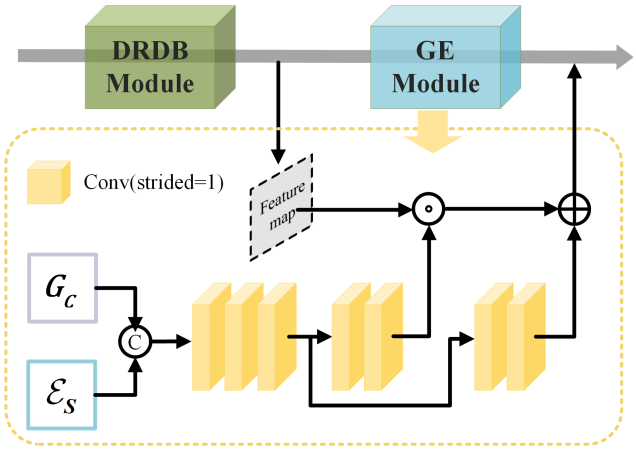

Figure 4. The implementation of our proposed gradient enhancement (GE) module for HDR image structure enhancement.

3.3. Optimization Strategy

Our proposed ERS-HDRI framework is optimized in a coarse-to-fine strategy, encompassing both the coarse HDRI optimization and fine HDRI optimization, where L_{coarse} is conducted to constrain the E-DRE network and L_{fine} is conducted to guide the G-HDRR network.

Coarse HDRI Optimization. Our event-based dynamic range enhancement network aims to learn the coarse HDRI function by adaptively fusing multi-modal data. Therefore, the coarse HDRI loss is conducted by employing L1 loss and perceptual loss [53] at both pixel and feature levels. Moreover, considering that the color information of LDR images is severely damaged and events contain less color information to compensate, the GAN loss is employed to perform effectively in image-coloring tasks to recover the color information and encourage the network to generate more natural images. The overall coarse HDRI loss $\mathcal{L}_{coarse}$ is formulated as follows:

$$\begin{aligned}\mathcal{L}_{coarse} &= \|\hat{I}_{coarse} - I\|_1 \\ &+ \lambda_{per} \sum_i \frac{\lambda_i}{C_i H_i W_i} \left\|\phi_i(\hat{I}_{coarse}) - \phi_i(I)\right\|_2^2 \\ &+ \mathbb{E}[-\log(D(\hat{I}_{coarse})] + \mathbb{E}[-\log(1 - D(I))],\end{aligned} \tag{8}$$

where I is the ground-truth HDR image, ϕ_i represents the i-th layer in pre-trained VGG-19 network [54], λ_i denotes the weight of the i-th feature map, C_i, H_i, W_i are the shapes of the feature map of the i-th layer, D represents the discriminator, and λ_{per} controls the trade-off between these three terms.

Fine HDRI Optimization. For the gradient-enhanced HDR reconstruction module, both spatial loss and reconstruction loss are conducted for optimization. The spatial loss calculates the difference of high-frequency spatial information between the generated fine HDR image and the ground-truth image, which fascinates the structure enhancement process. For the reconstruction loss, the multiscale fusion module is optimized by L1, VGG, and GAN losses similar to the coarse HDRI process; thus, the fine HDRI loss function is formulated as follows:

$$\begin{aligned}\mathcal{L}_{fine} &= \left\|\nabla AP\left(\hat{I}_{fine}\right) - \nabla AP(I)\right\|_2^2 \\ &+ \|\hat{I}_{fine} - I\|_1 \\ &+ \lambda_{per} \sum_i \frac{\lambda_i}{C_i H_i W_i} \left\|\phi_i(\hat{I}_{fine}) - \phi_i(I)\right\|_2^2 \\ &+ \mathbb{E}[-\log(D(\hat{I}_{fine})] + \mathbb{E}[-\log(1 - D(I))],\end{aligned} \tag{9}$$

where ∇ denotes the gradient operator employed for extracting high-frequency spatial information and $AP(\cdot)$ represents the average pooling function applied along the channel dimension.

Finally, the overall loss function for our ERS-HDRI framework is defined as follows:

$$\mathcal{L}_{total} = \lambda_{coarse}\mathcal{L}_{coarse} + \lambda_{fine}\mathcal{L}_{fine}. \tag{10}$$

where λ_{coarse} and λ_{fine} are the hyperparameters that control the trade-off of each term.

4. ERS-HDRD Dataset

Due to the lack of available datasets for the effective evaluation of our proposed method, both real-world and synthetic datasets are constructed, and the designs and processes are elaborated on.

4.1. Real-World Dataset

In this subsection, the designed hybrid camera system for capturing the ERS-HDRD dataset in real-world scenes is first introduced, followed by the presentation of the dataset details.

4.1.1. Hybrid Camera System

As shown in Figure 5, the hybrid camera system is equipped with two different cameras, an RGB camera, i.e., FLIR BFS-U3-32S4, for capturing real LDR images at 10 FPS, and an event camera, i.e., Prophesee Gen 4.1 EVK2-HD camera, for collecting concurrent event streams, fixed to the component. The dimension design of the hybrid imaging system takes into account two primary considerations. Firstly, in light of the constraints posed by the aircraft's observation window, it is imperative that the overall structure of the hybrid imaging system remains within specified size limitations. Secondly, to keep spatial calibration accuracy and preserve image and event resolution after spatial calibration, it is essential to minimize the baseline (distances between two cameras). Therefore, the dimensions of the component are set to 270 mm $\times$ 174 mm $\times$ 112 mm, with the centers of the apertures for the two cameras positioned at a separation distance of 57 mm. Temporal synchronization and spatial calibration are conducted as follows.

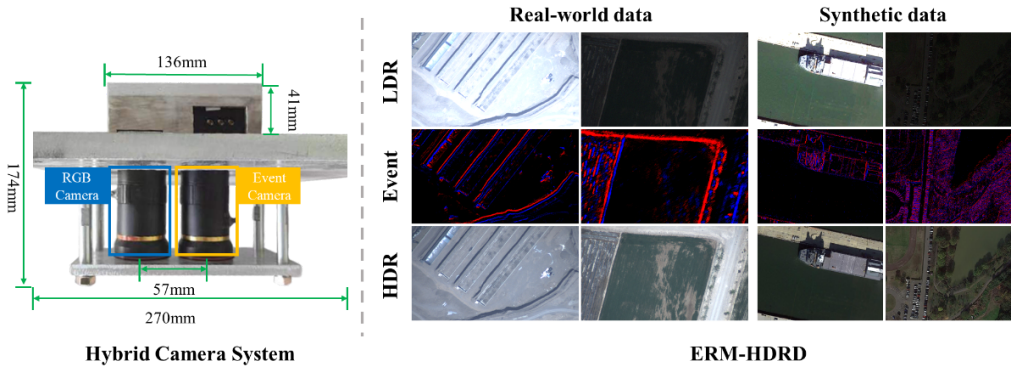

Figure 5. The details of our proposed datasets. (**Left**): the hardware implementation of our hybrid camera system. (**Right**): samples from our proposed ERS-HDRD dataset, composed of both real-world and synthetic data.

Temporal Synchronization. The two cameras are synchronized by using the trigger signal generated by the RGB camera. Specifically, the RGB camera generates alternating positive and negative triggers at the beginning and end of exposures. The event camera captures the signal from the RGB camera, recording accurate timestamps and polar information of the triggers. This methodology enables the extraction of events precisely aligned with the exposure time of LDR images through the recorded triggers.

Spatial Calibration. Spatial calibration is performed between the RGB camera and the event camera to guarantee a consistent field of view across all cameras. The distinct modalities of information output by the two cameras pose challenges to direct spatial alignment. To address this complexity, the E2VID methodology [34] is employed for reconstructing the intensity map of the event. Subsequently, the alignment between events and frames is performed by using the transformation matrix estimated by matching SIFT features [55] between the frame captured by the RGB camera and the reconstructed intensity maps. Specifically, perspective transformation [56] is applied to the RGB images as shown in Equation (11).

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = A \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \tag{11}$$

where A is the transformation matrix, (x, y, z) and (x', y', z') denote the source and destination image pixels.

To demonstrate the performance of our spatial calibration strategy, we compared it with the affine transformation and handcrafted-perspective transformation, where the feature points are selected manually. The comparison results are presented in Figure 6. It can be observed that affine transformation, limited to linear alterations such as translation, scaling, and rotation, cannot meet the requirements of calibration between LDR images and events, as shown in Figure 6b. In contrast, perspective transformations, encompassing the third dimension in comparison to affine transformations, exhibit superior efficacy, as shown in Figure 6c,d. In Figure 6c, manual feature points are employed for performing handcrafted perspective transformations on the original LDR image. This process is iteratively executed 10 times, each time selecting 20 corresponding points and retaining the best result. However, the effectiveness of this method depends on the accuracy of the selected points, making it less robust. Notably, a discernible misalignment with events appears on the right edge in Figure 6c. Our method utilizes the SIFT algorithm to detect and match features between the LDR image and event intensity map at various scales, extracting a large number of key points, thus yielding high stability and optimal results, as illustrated in Figure 6d.

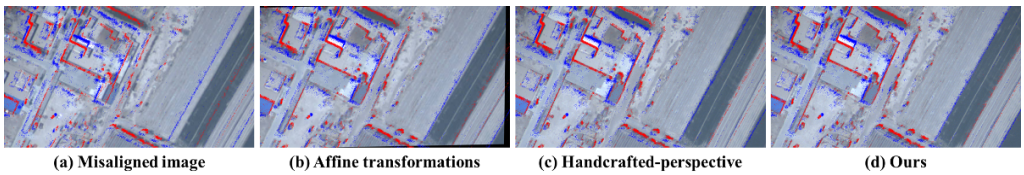

Figure 6. Alignment of LDR images with events. The events (red is positive, and blue is negative) are overlain with the LDR image.

4.1.2. Dataset Setup

The comprehensive system is affixed to the aircraft, capturing images at an altitude of approximately 5000 m under diverse lighting conditions. Moreover, 20,000 alternatively exposed LDR images and their corresponding event streams are captured in diverse scenes, such as lakes, mountains, buildings, and deserts, with a resolution of 1280 × 720. By aligning and merging the alternatively exposed LDR images following [57], the ground-truth HDR reference can be obtained. The training and testing parts are divided into a ratio of 9:1. Note that the LDR images suffer from both LDR and noise when captured in low-light conditions.

4.2. Synthetic Dataset

The synthetic dataset is built upon the remote sensing dataset DOTA-v1.0 [1], which contains a large amount of RGB images with resolutions ranging from 800 × 800 to 20,000 × 20,000 pixels. These images encompass a diverse of object categories and, thus, can also be used for external verification on the object detection task. By carrying out pre-

processing that involves the removal of samples with extreme aspect ratios, 2340 images are obtained as the ground-truth reference. Subsequently, a sliding window methodology is employed for each image, yielding the generation of 600 frames with randomized strides. The ESIM simulator [58] is utilized to simulate corresponding events and the frame rate is configured to 100. Following this, a series of LDR images with extreme exposure is generated through the brightness transform [59] of the original HDR images. To simulate the presence of noise in under-exposed images in real-world scenarios, we further process the under-exposed images by introducing Gaussian noise following the previous works [60–62]. The noise level dynamically increases as the lighting conditions decrease and varies between 0 and 50 in response to changes in light conditions. Ultimately, 38,400 data pairs are selected as the synthetic dataset, with a resolution of 640 × 480. Moreover, 34,560 samples are adopted as the training data and the remaining 3840 samples are adopted as the testing part.

4.3. Comparison with Existing HDRI Datasets

To contextualize the ERS-HDRD dataset within the broader landscape of HDRI and remote sensing research, statistical data about the existing HDRI datasets and our proposed ERS-HDRD dataset are presented in Table 1. Most works have focusd on conducting an HDRI benchmark for scenarios on the ground only based on the conventional camera [63,64]. Zhang et al. [5] explored the HDR reconstruction of low-illumination on remote sensing images and introduced the synthetic degeneration images from the VHR-10 dataset. However, the size of the dataset was too limited. Han et al. [19] constructed an event-based HDRI dataset named HES-HDR. Although its image resolution was high, the low resolution of the event camera limited the potential for HDRI. In contrast, our proposed ERS-HDRD dataset, which features high resolution in both images and events, focuses on challenges unique to the remote sensing field. Additionally, the ERS-HDRD dataset includes a large-scale dataset generated based on the DOTA-v1.0 [1] dataset, providing substantial dataset support for event-based HDRI.

Table 1. Comparison of different HDRI datasets. The remote option and event option respectively denote whether the dataset is constructed under remote sensing scenarios or by event cameras.

Dataset	Data Pairs	Remote	Event	Resolution (Image, Event)
Kalantari13 [63]	976	×	×	1280 × 720, NA
HDM-HDR-2014 [64]	15,087	×	×	1920 × 1080, NA
VHR-10-LI [5]	650	✓	×	1100 × 1100, NA
HES-HDR [19]	3071	×	✓	2448 × 2048, 346 × 260
ERS-HDRD (Real-world)	20,000	✓	✓	1280 × 720, 1280 × 720
ERS-HDRD (Synthetic)	38,400	✓	✓	640 × 480, 640 × 480

5. Experiments

This section compares our proposed approach with existing state-of-the-art methods on ERS-HDRD. First, the experimental settings, including the compared methods, metrics, and implementation details are introduced in Section 5.1. After that, quantitative and qualitative evaluations are provided on both the synthetic and real-world data and the results are analyzed in Section 5.2. Then, the external verification on the object detection task is carried out in Section 5.3, followed by the efficiency evaluation in terms of parameters and runtime performance in Section 5.4. Finally, comprehensive ablation experiments are conducted in Section 5.5 to verify the effectiveness of individual components of ERS-HDRI.

5.1. Experimental Settings

5.1.1. Comparison Methods and Metrics

This work compares the proposed ERS-HDRI with existing state-of-the-art methods, including the frame-based method DeepHDR [26], HDRUNet [11], KUNet [30], and the event-based method, HDRev [40]. To illustrate the overall performance of all the methods, various metrics are adopted for evaluation from different perspectives. We utilize the PSNR

and SSIM metrics to quantify the difference between the reconstructed images and the ground truth (GT) HDR reference. Moreover, to better simulate human visual perception, the learned perceptual image patch similarity (LPIPS) is employed for further evaluation.

5.1.2. Implementation Details

Our model is implemented in PyTorch and uses the ADAM [65] optimizer during the training process. It is trained on two NVIDIA GeForce RTX 4090 GPUs for 200 epochs. The initial learning rate is set to 0.0002 and decayed linearly from 100 epochs to the end. The hyperparameters $\{\lambda_{per}, \lambda_{coarse}, \lambda_{fine}\}$ are set as follows: {10, 1, 2}.

5.2. Comparison with State-of-the-Art Methods

5.2.1. Results on Synthetic Data

Comparisons with state-of-the-art methods are first conducted on the synthetic test data both quantitatively and qualitatively.

Quantitative Evaluation. As shown in Table 2, our method outperforms all the competitors in all metrics, demonstrating our algorithm's ability to reconstruct the HDR remote sensing image with different exposures. Specifically, DeepHDR [26], HDRUNet [11], and KUNet [30] only rely on a single LDR image with limited information, thus obtaining unsatisfactory results. Despite HDRev [40] also incorporating event and RGB images as inputs, it fails to integrate the information of the two modalities due to the distribution differences between remote sensing events and ground events, leading to missing details and displeasing color information. In contrast, our proposed approach takes the coarse-to-fine strategy and leverages the intra- and cross-attention (ICA) module to realize efficient and adaptive fusion between the LDR frame and high dynamic range events, thus reconstructing more natural and informative results and obtaining the PSNR gain of up to 12.263 dB and the LPIPS gain of up to 0.174. In addition, by leveraging the gradient enhancement (GE) module, our method significantly improves the sharpness of the reconstructed results, thus obtaining a significant performance improvement on SSIM compared with the second-best approach (HDRUNet), i.e., 32.63%.

Table 2. Quantitative comparisons on the synthetic data. **Bold** and underlined numbers represent the best and second-best performances. The symbol ↑ indicates that the higher the value, the better, and the symbol ↓ indicates that the lower the value, the better.

Metrics	Frame-Based Methods			Event-Based Methods	
	DeepHDR [26]	HDRUnet [11]	KUnet [30]	HDRev [40]	Ours
PSNR ↑	<u>16.865</u>	11.837	11.448	12.766	**29.128**
SSIM ↑	0.627	<u>0.668</u>	0.659	0.560	**0.886**
LPIPS ↓	0.293	<u>0.229</u>	0.242	0.362	**0.055**

Qualitative Evaluation. The visual comparisons of results on an over-exposed LDR image are presented in Figure 7. It can be seen that our approach addresses the challenge of over-exposed areas by restoring intricate scene textures and presenting color information more aligned with human perception. In contrast, other methods exhibit notable artifacts and color blocks, e.g., the trees and ground in the green box. This problem is attributed to the deficiency of texture information in saturated areas of RGB images, rendering the recovery of HDR images from a single RGB image ill-posed. To overcome this limitation, our ERS-HDRI incorporates event information from another modality, augmenting existing methodologies and yielding superior results. Notably, through the introduction of gradient feature extraction and reinforcement module, our method enhances the depiction of texture, e.g., the white marker lines on the playground within the delineated red box.

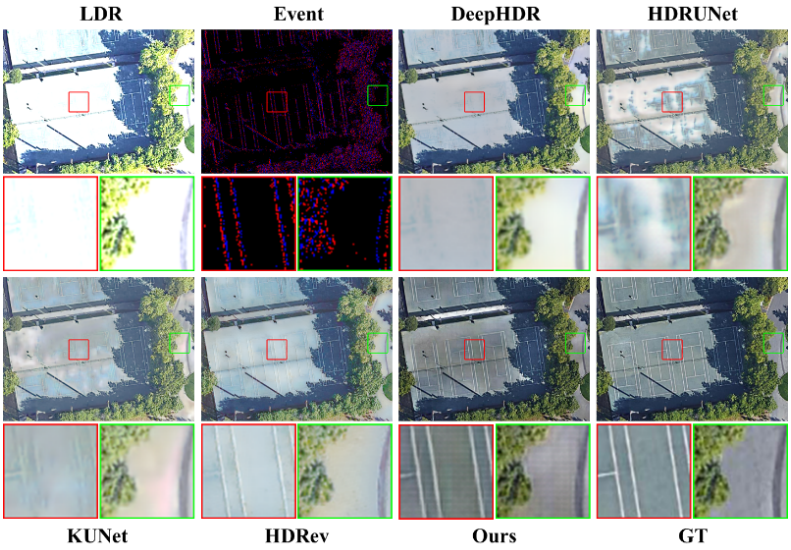

Figure 7. Qualitative results on over-exposed images in synthetic data. GT represents the ground-truth HDR reference.

Simultaneously, this subsection presents the outcomes in low-light conditions, as illustrated in Figure 8. It can be found that all methodologies exhibit a general capacity to restore the subject content of the image, which is attributed to the preservation of certain texture information in low-light conditions. However, a notable observation is the prevalence of considerable noise in the region restored by DeepHDR [26], e.g., the detail in the green box. HDRUet [11] and KUNet [30] achieve HDR image reconstruction by globally increasing brightness, leading to instances of local overexposure, e.g., the harbor within the red box. Moreover, HDRev [40] tends to generate false and aesthetically displeasing color information, e.g., the buildings in the green box. In comparison, our proposed method exploits the high dynamic range of events, along with the color information from RGB images, generating HDR images that exhibit enhanced visual appeal, informativeness, and clarity. Moreover, thanks to the multiscale denoising module introduced into the E-DRE network, our method shows excellent performance in noise suppression.

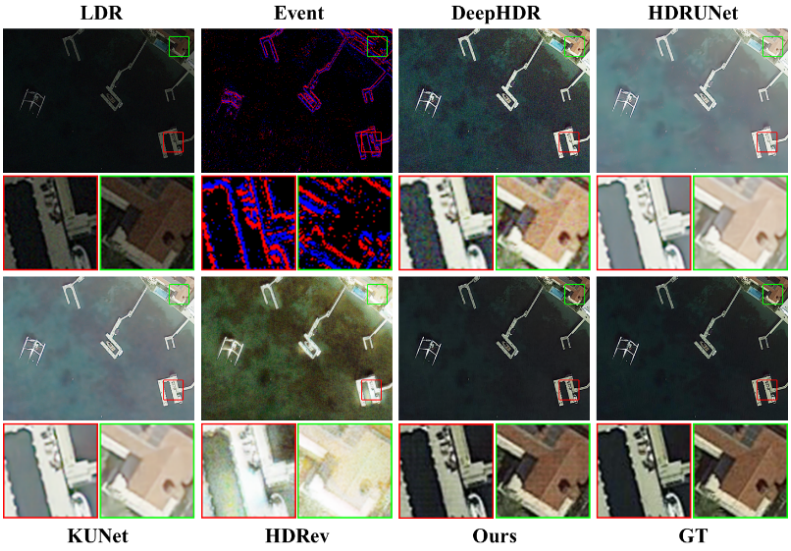

Figure 8. Qualitative results on under-exposed images in synthetic data. GT represents the ground-truth HDR reference.

5.2.2. Results on Real-World Data

In contrast to synthetic data, the task of HDRI in real scenes with remote sensing data poses increased challenges, which are attributed to the intricate nature of the scenes and the occurrence of mixing distortions. Consequently, experiments on real-world data are conducted to assess the robustness of our model.

Quantitative Evaluation. The overall comparisons are presented in Table 3. As can be observed, our proposed ERS-HDRI achieves superior performance in terms of all metrics from diverse perspectives, on average 6.043 gain on PSNR, 0.064 on SSIM, and 0.162 on LPIPS. The results show that the similarity and correlation between the HDRI reconstructed by ERS-HDRI and ground-truth images are higher and our reconstructed images enjoy less distortion and larger gradient amplitude. This achievement is attributed to the efficacy of our coarse-to-fine framework, encompassing the joint multi-modal fusion module and structure enhancement module. This integrated approach facilitates the restoration of badly exposed regions while preserving the original texture information to a large degree. Furthermore, these results demonstrate the robust feature extraction capabilities of our network, particularly when confronted with complex scenes.

Qualitative Evaluation. Firstly, the efficacy of our model in addressing overexposure is evaluated, as illustrated in Figure 9. Notably, our method clearly captures texture details across diverse scenes, e.g., the delineation of mountain peak structures within the red box, while other methodologies exhibit extensive color patches and blurring, emphasizing the value of the supplementary information provided by events. Although other methods perform well in restoring original information for mildly over-exposed regions, the frame-based method DeepHDR [26], HDRUNet [11], and KUNet [30] struggle to restore edge information for severely over-exposed areas, leading to pronounced blurring. Although event-based HDRev [40] achieves commendable results closely approaching that of our model within certain scenes, the abnormal color mapping reduces the contrast of some targets, thereby compromising visual effectiveness, as shown in the example. This deficiency can be attributed to HDRev's inadequate consideration of the distinctions between the two modal data during the integration of event information. In contrast, our method strategically addresses this issue through a feature alignment process, effectively narrowing the domain gap between multi-modalities.

In real scenes, under-exposed images frequently exhibit substantial noise, which leads to artifacts in the reconstructed HDR. The comparative visualization of various methods on under-exposed images is presented in Figure 10. Notably, both DeepHDR [26] and HDRev [40] encounter considerable noise, resulting in the decay and loss of intricate textures, e.g., the detail in the red box. In contrast, our proposed ERS-HDRI demonstrates superior noise handling capabilities, yielding clear HDR images that closely align with ground-truth representations. HDRUet [11] and KUNet [30], although enhancing overall brightness, impose stringent constraints on pixel brightness values within a narrow range, leading to a significant reduction in contrast and the diminishing of target characteristics. Leveraging a double-branch structure in our approach allows us not only to recover lost information in under-exposed areas but also to enhance the texture of the restored results and mitigate the impact of noise.

Table 3. Quantitative comparisons on real-world data. **Bold** and underlined numbers represent the best and second-best performances. The symbol ↑ indicates that the higher the value, the better, and the symbol ↓ indicates that the lower the value, the better.

Metrics	Frame-Based Methods			Event-Based Methods	
	DeepHDR [26]	HDRUnet [11]	KUNet [30]	HDRev [40]	Ours
PSNR ↑	<u>20.183</u>	17.693	17.514	16.368	**26.226**
SSIM ↑	0.678	0.716	<u>0.728</u>	0.540	**0.792**
LPIPS ↓	0.284	<u>0.273</u>	0.278	0.471	**0.111**

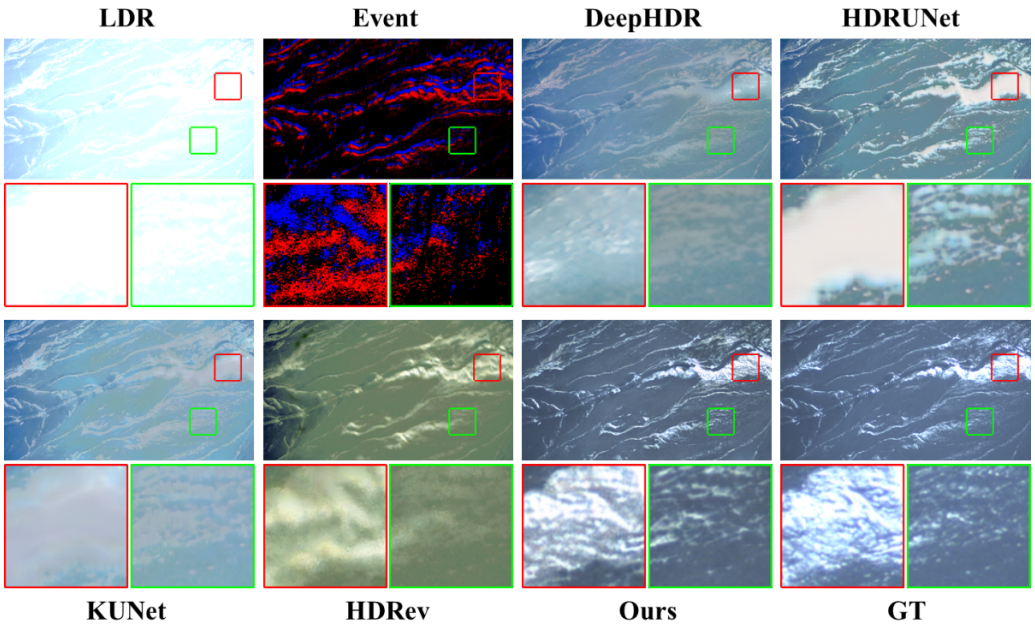

Figure 9. Qualitative results on over-exposed images in real-world data. GT represents the ground-truth HDR reference.

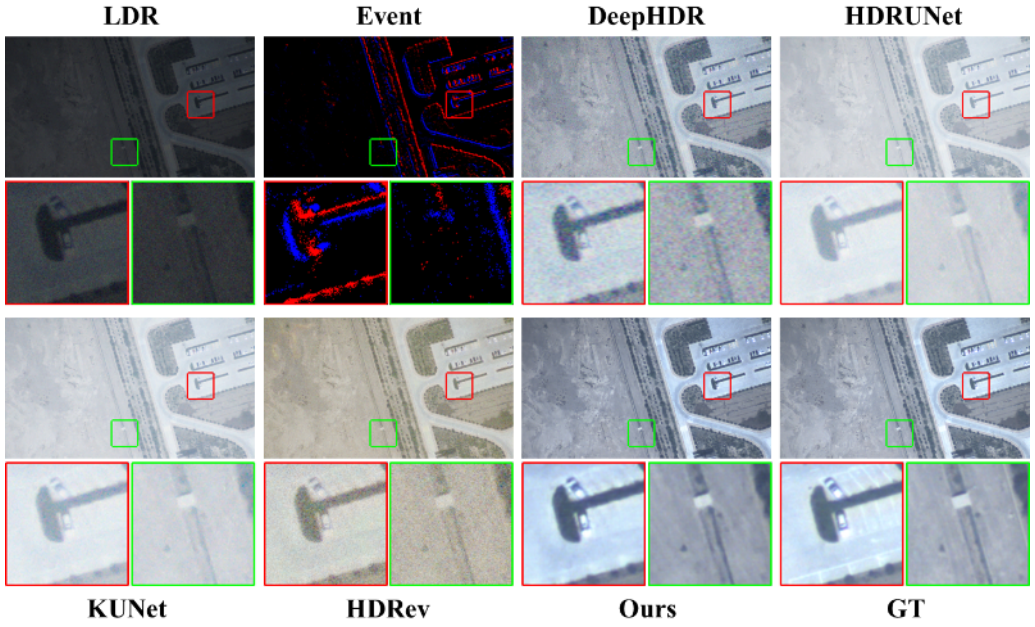

Figure 10. Qualitative results on under-exposed images in real-world data. GT represents the ground-truth HDR reference.

5.3. External Verification on Object Detection

To assess the efficacy of our proposed method in enhancing downstream tasks, various HDRI methods are taken as pre-processing steps for object detection, subsequently comparing the detection results. Particularly, LSKNet, introduced by Li et al. [66], is employed as the detector, given its recent advancements in utilizing large and selective kernels for remote sensing object detection, leading to state-of-the-art performance on competitive benchmarks such as DOTA. Specifically, the LSKNet-T version, fine-tuned on the DOTA for our detection experiments, is adopted in the experiment. The visual detection results are presented in Figures 11 and 12, indicating a substantial enhancement in detection efficacy after applying our method, irrespective of over-exposed or under-exposed scenes. The

improved detection is particularly evident in the identification of smaller targets, e.g., the detail in the red box.

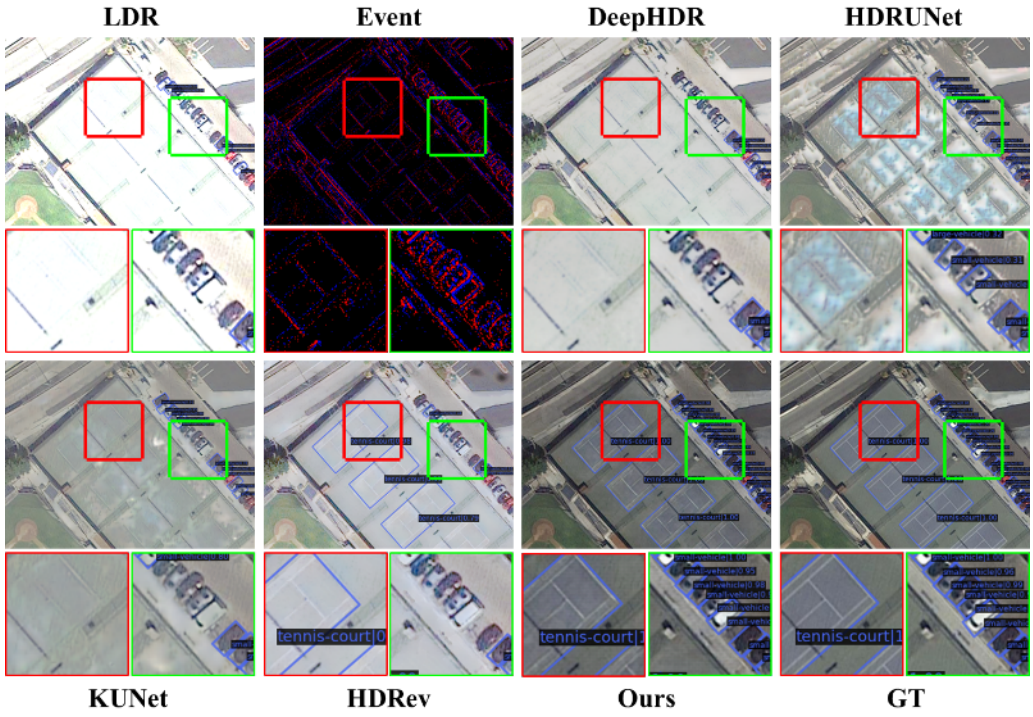

Figure 11. Detection results of LSKNet [66] on over-exposed LDR images processed by different methods. GT represents the ground-truth HDR reference.

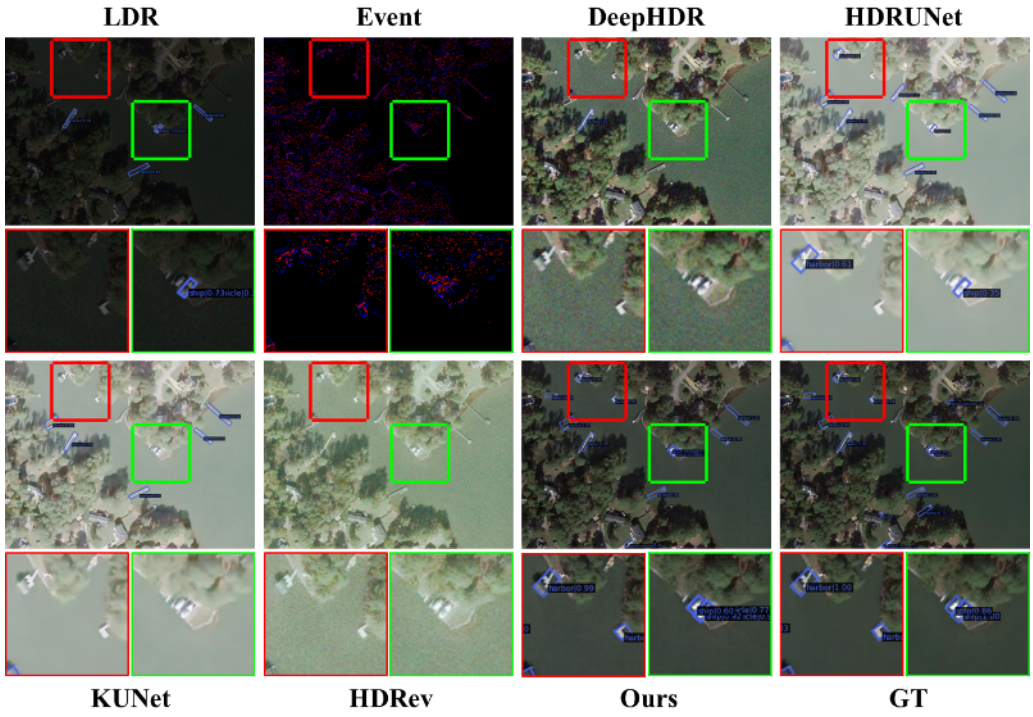

Figure 12. Detection results of LSKNet [66] on under-exposed LDR images processed by different methods. GT represents the ground-truth HDR reference.

5.4. Efficiency Evaluation

To demonstrate the efficiency of our ERS-HDRI, we evaluate the HDRI methods in different dimensions, including the amounts of parameters (Param) to compare the size of models and floating point operations (FLOPs) as well as runtime to evaluate the speed. Specifically, we implement all methods on an NVIDIA RTX 4090 GPU and obtain the runtime by averaging the cost time of 3840 images in our ERS-HDRD dataset with a size of 640 × 480. As shown in Table 4, although DeepHDR shows its superiority on the FLOPs and runtime, the heavy parameters limit its application on the edge device. In contrast, HDRUnet, KUnet, and our proposed method obtain more lightweight models, while our methods achieve better performance in inference time.

Table 4. Efficiency evaluation with other HDRI methods on the parameter, FLOPs, and runtime. **Bold** and underlined numbers represent the best and second-best performances.

	DeepHDR [26]	**HDRUnet [11]**	**KUnet [30]**	**HDRev [40]**	**Ours**
# Param (M)	51.54	<u>1.65</u>	**1.14**	57.93	10.06
FLOPs (10^9)	**75.77**	203.11	217.66	748.47	<u>194.41</u>
Runtime (s)	**0.0161**	0.0215	0.0273	0.4661	<u>0.0209</u>

5.5. Ablation Study

In this subsection, ablation studies are conducted to prove the effectiveness of our key modules in the ERS-HDRI framework. A baseline model is first trained by utilizing the E-DRE network with only remote sensing LDR images as input and then the ablation study is implemented by incrementally adding them over the baseline. Ablation studies are conducted as follows. (a) **baseline**: the baseline framework; (b) **+event**: adding event streams into the framework to compensate for the HDRI process; (c) **+ICA**: further adopting intra- and cross-attention module into the E-DRE network. (d) **+G-HDRR**: further adding the G-HDRR into the ERS-HDRI framework to enhance the structure. Quantitative results in terms of PSNR [11], SSIM [27], and LPIPS on ERS-HDRD are presented in Table 5 and the best and second-best results are indicated in bold and underlined fonts, respectively. Meanwhile, the corresponding qualitative ablation results are also shown in Figure 13.

Importance of Events. Event streams contain high dynamic range information that can compensate for the HDRI process. By introducing them into the framework, one can improve the recovery capability of the network and generate a more informative remote sensing image. As shown in Figure 13, the results generated by fusing multi-modal data show the details in the over-exposed regions. Also, the quantitative results shown in Table 5 demonstrate the improvements of introducing event streams, achieving a PSNR gain of up to 2.421 dB, an SSIM gain of up to 0.034, and an LPIPS gain of up to 0.106.

Importance of ICA. To explore the impact of the ICA module on the ERS-HDRI performance, the proposed ERS-HDRI is evaluated by replacing the concatenation operation with the ICA module to improve the efficacy of the multi-modal fusion process. As shown in Figure 13, with the help of ICA, the reconstructed HDR image is more visually pleasant and suffers less distortion, e.g., the tiles on the roof. From Table 5, it can be seen that the PSNR value of the ICA-based network is 2.129 dB higher than that of the concatenation-based network. Moreover, the SSIM and LPIPS performances are increased again by 0.024 and 0.025.

Importance of G-HDRR. To verify the effectiveness of the proposed G-HDRR network, we perform the cascading process of the E-DRE network and G-HDRR network. As demonstrated in Table 5, with the combination of the G-HDRR network, our method further improves the performance metrics, i.e., 0.435 dB PSNR gain, 0.005 SSIM gain, and 0.013 LPIPS gain. The reconstructed HDR result is also sharper and clearer in the low-contrast regions, as shown in Figure 13.

Table 5. Comparisons of average PSNR (dB) and SSIM on our real-world dataset in ablation settings. **Bold** and underlined numbers represent the best and second-best performances. The symbol ↑ indicates that the higher the value, the better, and the symbol ↓ indicates that the lower the value, the better.

Baseline	+Events	+ICA	+G-HDRR	PSNR ↑	SSIM ↑	LPIPS ↓
✓				21.241	0.729	0.255
✓	✓			23.662	0.763	0.149
✓	✓	✓		25.791	0.787	0.124
✓	✓	✓	✓	**26.226**	**0.792**	**0.111**

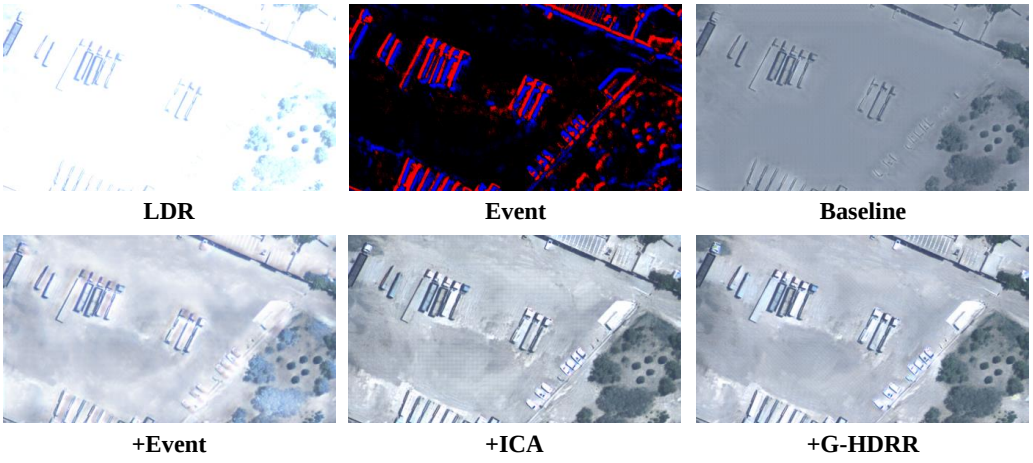

Figure 13. Visual comparisons on the real-world dataset of ERS-HDRD in ablation settings.

6. Conclusions and Discussion

In this paper, an event-based remote sensing HDR imaging framework named ERS-HDRI is proposed to handle the limitation of the conventional camera, which can reconstruct the HDR remote sensing image from a single-exposure LDR image and its concurrent event streams by leveraging a coarse-to-fine strategy. Specifically, the event-based dynamic range enhancement (E-DRE) network is designed to first extract the dynamic range features from LDR frames and events and then perform multi-modal feature fusion adaptively with the intra- and cross-attention modules. To reduce the noise and generate more informative results, the multiscale denoise network and dense feature fusion network are then introduced to reconstruct the coarse clean HDR image. To enhance the missing information caused by light attenuation, our proposed framework further builds the gradient-enhanced HDR reconstruction (G-HDRR) network upon the gradient enhancement module and dense residual blocks to reconstruct the detailed gradient information in low-contrast regions. A novel remote sensing event-based HDRI dataset, i.e., ERS-HDRD, is also conducted for evaluation, which contains aligned LDR images, event streams, and corresponding HDR images. Experiments show that the proposed method outperforms the state-of-the-art methods on both synthetic data and real-world data.

Our method demonstrates the compensation ability of event cameras to traditional cameras in badly exposed remote sensing scenes, enhancing the ability to capture more informative remote sensing images, and can further improve the accuracy of subsequent tasks such as remote sensing object detection and tracking. Some similar remote sensing tasks, such as backlit image enhancement [67] and low-light image enhancement [5], can also be addressed with our framework. In addition, event cameras have other important attributes, such as high temporal resolution and low latency, which could be further developed to address challenges in other tasks, such as remote sensing motion deblurring [68], by leveraging our methods.

7. Limitations and Future Work

Although our ERS-HDRI demonstrates commendable efficacy in the remote sensing HDRI task, certain challenges persist in achieving precise color restoration when processing extremely over-exposed regions. The inherent limitation arises from the absence of color information in both low dynamic range (LDR) images and event streams, as shown in Figure 14. A prospective solution to this issue could involve the integration of image colorization techniques [69,70] as a post-processing step in future work.

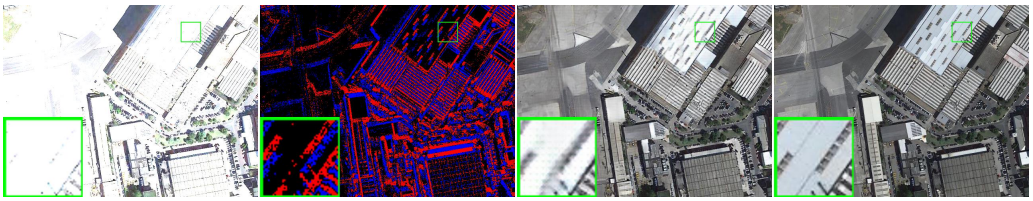

Figure 14. An example of a color reconstruction failure of our ERS-HDRI in the remote sensing HDRI process. GT represents the ground-truth HDR reference.

Author Contributions: Conceptualization, X.L.; methodology, X.L. and S.C.; validation, X.L. and S.C.; investigation, X.L., Z.Z. and S.C.; data curation, X.L., Z.Z. and S.C.; writing—original draft preparation, X.L. and S.C.; writing—review and editing, X.L., S.C., Z.Z., C.Z. and C.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets presented in this article are not readily available due to privacy.

Acknowledgments: The authors would like to thank Xia et al. for providing free DOTA-v1.0.

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

References

1. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 19–21 June 2018; pp. 3974–3983.
2. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [CrossRef]
3. Requena-Mesa, C.; Benson, V.; Reichstein, M.; Runge, J.; Denzler, J. EarthNet2021: A large-scale dataset and challenge for Earth surface forecasting as a guided video prediction task. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 19–25 June 2021; pp. 1132–1142.
4. Xiong, Z.; Zhang, F.; Wang, Y.; Shi, Y.; Zhu, X.X. Earthnets: Empowering ai in earth observation. *arXiv* **2022**, arXiv:2210.04936.
5. Zhang, X.; Zhang, L.; Wei, W.; Ding, C.; Zhang, Y. Dynamic Long-Short Range Structure Learning for Low-Illumination Remote Sensing Imagery HDR Reconstruction. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 859–862.
6. Xu, H.; Ma, J.; Le, Z.; Jiang, J.; Guo, X. Fusiondn: A unified densely connected network for image fusion. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12484–12491.
7. 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]
8. Lu, P.Y.; Huang, T.H.; Wu, M.S.; Cheng, Y.T.; Chuang, Y.Y. High dynamic range image reconstruction from hand-held cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 509–516.
9. Chen, Y.; Jiang, G.; Yu, M.; Yang, Y.; Ho, Y.S. Learning stereo high dynamic range imaging from a pair of cameras with different exposure parameters. *IEEE Trans. Comput. Imaging* **2020**, *6*, 1044–1058. [CrossRef]
10. Lee, S.H.; Chung, H.; Cho, N.I. Exposure-structure blending network for high dynamic range imaging of dynamic scenes. *IEEE Access* **2020**, *8*, 117428–117438. [CrossRef]
11. Chen, X.; Liu, Y.; Zhang, Z.; Qiao, Y.; Dong, C. Hdrunet: Single image hdr reconstruction with denoising and dequantization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 19–25 June 2021; pp. 354–363.
12. Kalantari, N.K.; Ramamoorthi, R. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.* **2017**, *36*, 144. [CrossRef]

13. Liu, Z.; Lin, W.; Li, X.; Rao, Q.; Jiang, T.; Han, M.; Fan, H.; Sun, J.; Liu, S. ADNet: Attention-guided deformable convolutional network for high dynamic range imaging. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 19–25 June 2021; pp. 463–470.
14. Vijay, C.S.; Paramanand, C.; Rajagopalan, A.N.; Chellappa, R. Non-uniform deblurring in HDR image reconstruction. *IEEE Trans. Image Process.* **2013**, *22*, 3739–3750. [CrossRef]
15. Lakshman, P. Combining deblurring and denoising for handheld HDR imaging in low light conditions. *Comput. Electr. Eng.* **2012**, *38*, 434–443. [CrossRef]
16. Gallego, G.; Delbrück, T.; Orchard, G.; Bartolozzi, C.; Taba, B.; Censi, A.; Leutenegger, S.; Davison, A.J.; Conradt, J.; Daniilidis, K.; et al. Event-based vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 154–180. [CrossRef]
17. Reinbacher, C.; Munda, G.; Pock, T. Real-time panoramic tracking for event cameras. In Proceedings of the IEEE International Conference on Computational Photography, Stanford, CA, USA, 12–14 May 2017; pp. 1–9.
18. Han, J.; Zhou, C.; Duan, P.; Tang, Y.; Xu, C.; Xu, C.; Huang, T.; Shi, B. Neuromorphic camera guided high dynamic range imaging. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1730–1739.
19. Han, J.; Yang, Y.; Duan, P.; Zhou, C.; Ma, L.; Xu, C.; Huang, T.; Sato, I.; Shi, B. Hybrid high dynamic range imaging fusing neuromorphic and conventional images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 8553–8565. [CrossRef]
20. Messikommer, N.; Georgoulis, S.; Gehrig, D.; Tulyakov, S.; Erbach, J.; Bochicchio, A.; Li, Y.; Scaramuzza, D. Multi-Bracket High Dynamic Range Imaging with Event Cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 547–557.
21. Akyüz, A.O.; Fleming, R.; Riecke, B.E.; Reinhard, E.; Bülthoff, H.H. Do HDR displays support LDR content? A psychophysical evaluation. *ACM Trans. Graph. (TOG)* **2007**, *26*, 38-es. [CrossRef]
22. Masia, B.; Agustin, S.; Fleming, R.W.; Sorkine, O.; Gutierrez, D. Evaluation of reverse tone mapping through varying exposure conditions. In *ACM SIGGRAPH Asia 2009 Papers*; Pacifico Yokohama: Yokohama, Japan, 2009; pp. 1–8.
23. Kovaleski, R.P.; Oliveira, M.M. High-quality reverse tone mapping for a wide range of exposures. In Proceedings of the 2014 27th SIBGRAPI Conference on Graphics, Patterns and Images, Rio de Janeiro, Brazil, 26–30 August 2014; pp. 49–56.
24. Masia, B.; Serrano, A.; Gutierrez, D. Dynamic range expansion based on image statistics. *Multimed. Tools Appl.* **2017**, *76*, 631–648. [CrossRef]
25. Eilertsen, G.; Kronander, J.; Denes, G.; Mantiuk, R.K.; Unger, J. HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph.* **2017**, *36*, 1–15. [CrossRef]
26. Santos, M.S.; Ren, T.I.; Kalantari, N.K. Single image HDR reconstruction using a CNN with masked features and perceptual loss. *arXiv* **2020**, arXiv:2005.07335.
27. Liu, Y.L.; Lai, W.S.; Chen, Y.S.; Kao, Y.L.; Yang, M.H.; Chuang, Y.Y.; Huang, J.B. Single-image HDR reconstruction by learning to reverse the camera pipeline. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1651–1660.
28. Akhil, K.; Jiji, C. Single Image HDR Synthesis Using a Densely Connected Dilated ConvNet. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Online, 19–25 June 2021; pp. 526–531.
29. A Sharif, S.; Naqvi, R.A.; Biswas, M.; Kim, S. A two-stage deep network for high dynamic range image reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 19–25 June 2021; pp. 550–559.
30. Wang, H.; Ye, M.; Zhu, X.; Li, S.; Zhu, C.; Li, X. KUNet: Imaging Knowledge-Inspired Single HDR Image Reconstruction. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, Vienna, Austria, 23–29 July 2022; pp. 1408–1414.
31. Hu, X.; Shen, L.; Jiang, M.; Ma, R.; An, P. LA-HDR: Light Adaptive HDR Reconstruction Framework for Single LDR Image Considering Varied Light Conditions. *IEEE Trans. Multimed.* **2022**, *25*, 4814–4829. [CrossRef]
32. Belbachir, A.N.; Schraml, S.; Mayerhofer, M.; Hofstätter, M. A novel hdr depth camera for real-time 3d 360 panoramic vision. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 425–432.
33. Bardow, P.; Davison, A.J.; Leutenegger, S. Simultaneous optical flow and intensity estimation from an event camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 884–892.
34. Rebecq, H.; Ranftl, R.; Koltun, V.; Scaramuzza, D. Events-to-video: Bringing modern computer vision to event cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3857–3866.
35. Wang, L.; Ho, Y.S.; Yoon, K.J.; Mohammad Mostafavi, I.S. Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10081–10090.
36. Scheerlinck, C.; Rebecq, H.; Gehrig, D.; Barnes, N.; Mahony, R.; Scaramuzza, D. Fast image reconstruction with an event camera. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass Village, CO, USA, 1–5 March 2020; pp. 156–163.

37. Zou, Y.; Zheng, Y.; Takatani, T.; Fu, Y. Learning to reconstruct high speed and high dynamic range videos from events. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 19–25 June 2021; pp. 2024–2033.
38. Liang, Q.; Zheng, X.; Huang, K.; Zhang, Y.; Chen, J.; Tian, Y. Event-Diffusion: Event-Based Image Reconstruction and Restoration with Diffusion Models. In Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October–3 November 2023; pp. 3837–3846.
39. Shaw, R.; Catley-Chandar, S.; Leonardis, A.; Pérez-Pellitero, E. HDR reconstruction from bracketed exposures and events. *arXiv* **2022**, arXiv:2203.14825.
40. Yang, Y.; Han, J.; Liang, J.; Sato, I.; Shi, B. Learning event guided high dynamic range video reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 13924–13934.
41. Gao, T.; Niu, Q.; Zhang, J.; Chen, T.; Mei, S.; Jubair, A. Global to local: A scale-aware network for remote sensing object detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5615614. [CrossRef]
42. Zhou, X.; Zhou, L.; Gong, S.; Zhong, S.; Yan, W.; Huang, Y. Swin Transformer Embedding Dual-Stream for Semantic Segmentation of Remote Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *17*, 175–189. [CrossRef]
43. Roy, S.K.; Deria, A.; Hong, D.; Rasti, B.; Plaza, A.; Chanussot, J. Multimodal fusion transformer for remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5515620. [CrossRef]
44. Chen, H.; Qi, Z.; Shi, Z. Remote sensing image change detection with transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [CrossRef]
45. Cheng, G.; Si, Y.; Hong, H.; Yao, X.; Guo, L. Cross-scale feature fusion for object detection in optical remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 431–435. [CrossRef]
46. Li, J.; Chen, L.; Shen, J.; Xiao, X.; Liu, X.; Sun, X.; Wang, X.; Li, D. Improved neural network with spatial pyramid pooling and online datasets preprocessing for underwater target detection based on side scan sonar imagery. *Remote Sens.* **2023**, *15*, 440. [CrossRef]
47. Han, L.; Zhao, Y.; Lv, H.; Zhang, Y.; Liu, H.; Bi, G. Remote sensing image denoising based on deep and shallow feature fusion and attention mechanism. *Remote Sens.* **2022**, *14*, 1243. [CrossRef]
48. Huang, Z.; Zhu, Z.; Wang, Z.; Shi, Y.; Fang, H.; Zhang, Y. DGDNet: Deep Gradient Descent Network for Remotely Sensed Image Denoising. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
49. Wang, J.; Li, W.; Wang, Y.; Tao, R.; Du, Q. Representation-enhanced status replay network for multisource remote-sensing image classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, 1–13. [CrossRef] [PubMed]
50. Xi, M.; Li, J.; He, Z.; Yu, M.; Qin, F. NRN-RSSEG: A deep neural network model for combating label noise in semantic segmentation of remote sensing images. *Remote Sens.* **2022**, *15*, 108. [CrossRef]
51. Dong, H.; Pan, J.; Xiang, L.; Hu, Z.; Zhang, X.; Wang, F.; Yang, M.H. Multi-scale boosted dehazing network with dense feature fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2157–2167.
52. Kim, Y.; Soh, J.W.; Park, G.Y.; Cho, N.I. Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3482–3492.
53. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 694–711.
54. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
55. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
56. Bayraktar, E.; Basarkan, M.E.; Celebi, N. A low-cost UAV framework towards ornamental plant detection and counting in the wild. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 1–11. [CrossRef]
57. Chen, G.; Chen, C.; Guo, S.; Liang, Z.; Wong, K.Y.K.; Zhang, L. HDR video reconstruction: A coarse-to-fine network and a real-world benchmark dataset. In Proceedings of the International Conference on Computer Vision, Online, 11–17 October 2021; pp. 2502–2511.
58. Rebecq, H.; Gehrig, D.; Scaramuzza, D. ESIM: An open event camera simulator. In Proceedings of the Conference on Robot Learning, Zürich, Switzerland, 29–31 October 2018; pp. 969–982.
59. Ying, Z.; Li, G.; Ren, Y.; Wang, R.; Wang, W. A new image contrast enhancement algorithm using exposure fusion framework. In Proceedings of the International Conference on Computer Analysis of Images and Patterns, Ystad, Sweden, 22–24 August 2017; pp. 36–46.
60. Li, M.; Liu, J.; Yang, W.; Sun, X.; Guo, Z. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Trans. Image Process.* **2018**, *27*, 2828–2841. [CrossRef]
61. Dhara, S.K.; Sen, D. Exposedness-based noise-suppressing low-light image enhancement. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 3438–3451. [CrossRef]
62. Li, X.; Fan, C.; Zhao, C.; Zou, L.; Tian, S. NIRN: Self-supervised noisy image reconstruction network for real-world image denoising. *Appl. Intell.* **2022**, *52*, 16683–16700. [CrossRef]
63. Kalantari, N.K.; Shechtman, E.; Barnes, C.; Darabi, S.; Goldman, D.B.; Sen, P. Patch-based high dynamic range video. *ACM Trans. Graph.* **2013**, *32*, 202. [CrossRef]

64. Pérez-Pellitero, E.; Catley-Chandar, S.; Leonardis, A.; Timofte, R. NTIRE 2021 challenge on high dynamic range imaging: Dataset, methods and results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 19–25 June 2021; pp. 691–700.
65. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
66. Li, Y.; Hou, Q.; Zheng, Z.; Cheng, M.M.; Yang, J.; Li, X. Large Selective Kernel Network for Remote Sensing Object Detection. *arXiv* **2023**, arXiv:2303.09030.
67. Lv, X.; Zhang, S.; Liu, Q.; Xie, H.; Zhong, B.; Zhou, H. BacklitNet: A dataset and network for backlit image enhancement. *Comput. Vis. Image Underst.* **2022**, *218*, 103403. [CrossRef]
68. Fang, J.; Cao, X.; Wang, D.; Xu, S. Multitask learning mechanism for remote sensing image motion deblurring. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 2184–2193. [CrossRef]
69. Sheng, Z.; Shen, H.L.; Yao, B.; Zhang, H. Guided colorization using mono-color image pairs. *IEEE Trans. Image Process.* **2023**, *32*, 905–920. [CrossRef]
70. Kang, X.; Lin, X.; Zhang, K.; Hui, Z.; Xiang, W.; He, J.Y.; Li, X.; Ren, P.; Xie, X.; Timofte, R.; et al. NTIRE 2023 video colorization challenge. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 1570–1581.