remote sensing

MDPI

Article

SFDA-MEF: An Unsupervised Spacecraft Feature Deformable Alignment Network for Multi-Exposure Image Fusion

Qianwen Xiong, Xiaoyuan Ren, Huanyu Yin, Libing Jiang, Canyu Wang and Zhuang Wang *

College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China; xiongqianwen0920@nudt.edu.cn (Q.X.); renxiaoyuan10@nudt.edu.cn (X.R.); yinhuanyu0418@nudt.edu.cn (H.Y.); jianglibing@nudt.edu.cn (L.J.); wangcanyu20@nudt.edu.cn (C.W.)
* Correspondence: wangzhuang@nudt.edu.cn

Abstract: Optical image sequences of spacecraft acquired by space-based monocular cameras are typically imaged through exposure bracketing. The spacecraft feature deformable alignment network for multi-exposure image fusion (SFDA-MEF) aims to synthesize a High Dynamic Range (HDR) spacecraft image from a set of Low Dynamic Range (LDR) images with varying exposures. The HDR image contains details of the observed target in LDR images captured within a specific luminance range. The relative attitude of the spacecraft in the camera coordinate system undergoes continuous changes during the orbital rendezvous, which leads to a large proportion of moving pixels between adjacent frames. Concurrently, subsequent tasks of the In-Orbit Servicing (IOS) system, such as attitude estimation, are highly sensitive to variations in multi-view geometric relationships, which means that the fusion result should preserve the shape of the spacecraft with minimal distortion. However, traditional methods and unsupervised deep-learning methods always exhibit inherent limitations in dealing with complex overlapping regions. In addition, supervised methods are not suitable when ground truth data are scarce. Therefore, we propose an unsupervised learning framework for the multi-exposure fusion of optical spacecraft image sequences. We introduce a deformable convolution in the feature deformable alignment module and construct an alignment loss function to preserve its shape with minimal distortion. We also design a feature point extraction loss function to render our output more conducive to subsequent IOS tasks. Finally, we present a multi-exposure spacecraft image dataset. Subjective and objective experimental results validate the effectiveness of SFDA-MEF, especially in retaining the shape of the spacecraft.

Keywords: multi-exposure image fusion; spacecraft optical image; unsupervised learning; dynamic scenes

Academic Editors: Xiong Xu, Zhenming Peng and Rujin Zhao

Received: 3 November 2024
Revised: 21 December 2024
Accepted: 6 January 2025
Published: 8 January 2025

Citation: Xiong, Q.; Ren, X.; Yin, H.; Jiang, L.; Wang, C.; Wang, Z. SFDA-MEF: An Unsupervised Spacecraft Feature Deformable Alignment Network for Multi-Exposure Image Fusion. *Remote Sens.* **2025**, *17*, 199. https://doi.org/10.3390/rs17020199

1. Introduction

In recent years, with the gradual development of space resources, the importance of developing technologies of space-based optical In-Orbit Servicing (IOS) has been increasing. Among these technologies, space-based imaging based on monocular cameras plays a significant role in challenging scenarios such as active debris removal. Due to the limitations involved in adverse illumination and the relative motion between the imaging platform and the observed spacecraft, a specific photographic technique called exposure bracketing is widely adopted in the observation scenario during orbital rendezvous, during which the camera takes multiple shots with different exposure times. By repeating the process of exposure bracketing, a continuous sequence of multi-exposure images of the observed

spacecraft is obtained which can capture details in both the highlights and shadows of the same scene with high contrast.

Multi-Exposure Fusion (MEF) imaging is the primary approach to creating a High Dynamic Range (HDR) image from a series of Low Dynamic Range (LDR) images with varying exposures. MEF gathers details under different exposure conditions. The aim of MEF for spacecraft images is to generate high-quality HDR images of the spacecraft which are visually sharp in texture and perform well in subsequent tasks of the IOS system.

Optical image sequences of spacecraft are typically obtained during an orbital rendezvous between the observation platform and the observed target. The relative attitude of the spacecraft in the camera coordinate system captured between adjacent frames with varying exposure times is constantly changing. Existing research related to spacecraft image fusion deals with multi-illumination angles image fusion [1] or visible-thermal infrared images fusion [2] rather than multi-exposure image fusion. At the same time, multi-exposure image fusion methods adopted in static scenes do not apply to spacecraft images [3–7]. Moreover, other methods for dynamic scenes mainly aim to image natural landscapes, buildings, or figures in scenarios with a complex background and a relatively small proportion of moving pixels. However, during imaging of spacecraft, both the observation platform and the observed target are in a state of relative motion. This leads to a large proportion of moving pixels on the part of the spacecraft as its relative attitude changes. Hence, during the MEF in the IOS system for spacecraft images, we mainly focus on the moving observed target.

When dealing with spacecraft images, traditional dynamic methods based on eliminating motion pixels have difficulty handling complex overlapping regions of the target, which may lead to loss of valid information [8,9]. Furthermore, some subsequent tasks in the IOS system possess certain particularities, such as three-dimensional reconstruction and attitude estimation. These are highly sensitive to variations in the multi-view geometric relationships of the observed target, while traditional registration-based methods are prone to causing changes in the shape of the observed targets (the corner of the solar panel, etc.) [10–12].

Methods based on deep learning have emerged as a new approach for MEF in dynamic scenes. However, the majority of these employ supervised models and require true HDR images as guidance [13–17]. Unsupervised methods based on GANs [18] can avoid this problem, but may give rise to false textures during reconstruction. Other unsupervised methods based on CNNs [19–22] exhibit limitations in effectively managing the large number of overlapping motion pixels, as they directly add feature maps when fusing and reconstructing the images. These limitations mean that they cannot be applied in cases where there are complex overlapping regions of spacecraft between adjacent frames. Meanwhile, for images in existing multi-exposure image fusion datasets, the cameras and still backgrounds are in a relatively static state, which is highly distinct from the spacecraft observation scenario. Moreover, existing open-source spacecraft datasets are generated solely based on a single exposure intensity. Neither of these situations can fulfill the practical application requirements for multi-exposure image fusion in an IOS system dealing with spacecraft images' lack of ground truth.

To address these issues, in this paper we propose a feature deformable alignment network based on unsupervised learning that applies to spacecraft multi-exposure image fusion for dynamic observation scenes, which we call SFDA-MEF. Our method can efficiently reconstruct the HDR image in the dynamic scene while effectively preserving the original multi-view geometric relationships of the observation targets. The main contributions of this paper include:

- An unsupervised learning framework is constructed to adopt multi-exposure fusion for optical images of spacecraft acquired by space-based monocular cameras.

- A feature deformable alignment module and corresponding alignment loss function are proposed. The deformable convolution is introduced to fit the complex geometric deformation of the observation spacecraft, which aids in extracting details of the target under varying exposures while preserving its shape with minimal distortion in the HDR image and maintaining consistency with the reference LDR image.
- A loss function based on the SIFT operator is designed to optimize the fusion results of the network, which makes it perform well in feature extraction and is more suitable for subsequent tasks such as attitude estimation within the IOS system.
- A new multi-exposure spacecraft image dataset is proposed to simulate imaging of spacecraft during orbital rendezvous.

2. Related Work

When examining the existing methods related to spacecraft image fusion, only two papers were found. Ao Xiang et al. [1] proposed a supervised network designed for the removal of shadows in spacecraft images and presented a multi-angle illumination spacecraft image dataset. Civardi et al. [2] implemented a realistic tool for rendering thermal infrared images of spacecraft targets and assessed different pixel-level techniques for fusing visible and thermal infrared images through qualitative and quantitative performance metrics. Consequently, when researching MEF methods for spacecraft images, our focus is restricted to current multi-exposure image fusion techniques and general image fusion methods that can be applied to multi-exposure image fusion.

2.1. Traditional Multi-Exposure Image Fusion Methods in Dynamic Scenes

Traditional MEF methods in dynamic scenes can be roughly categorized into two types based on their core concepts, namely, those relying on motion detection and those relying on image registration.

Methods based on motion detection usually select one of the LDR images as the reference image, then obtain the HDR image by marking and discarding the motion pixels in the non-reference images. Gallo et al. [8] predicted the exposure values of other images through the reference image. If the pixels deviating from the predicted exposure value in the actual image block exceed a certain proportion, it is determined that there are moving objects in this image area. Wang et al. [9] proposed a ghost elimination method based on stack expansion and visual saliency. The influence of the jumpiness of grayscale changes on fusion can be mitigated through stack expansion; at the same time, the information of the ghost regions is not involved in image fusion, helping to eliminate ghosts. Nevertheless, such methods are only applicable to scenarios with a small number of motion pixels. Spacecraft image sequences undergo change in the orientation of the observation target between adjacent frames, meaning that there are many moving pixels. This can lead to the loss of details or the phenomenon of discontinuous changes in pixel intensity in the generated HDR image.

On the other hand, methods based on image registration obtain a set of registered LDR images by using the alignment algorithm in the local or global range and then synthesizing the HDR images. Traditional optical flow methods are only applicable to the alignment of image sequences with consistent brightness. For example, Zimmer et al. [10] proposed an energy-based optical flow method for aligning multi-exposure sequences by minimizing an energy function that includes gradient terms and smoothness terms to ensure smooth reconstruction in saturated areas and then using the calculated displacement maps along with another energy function to reconstruct the HDR image. Because optical flow methods struggle to accurately estimate motion changes, Sen et al. [11] introduced an optimization scheme based on image blocks, integrating the generation of aligned image sequences and

multi-exposure image fusion into a single optimization problem. Hu et al. [12] transformed dynamic image sequences into static image sequences by simultaneously optimizing the energy function along with the continuity of both color and gradients prior to performing fusion. Although optimization-based methods can provide a superior fusion effect compared to optical flow-based methods, they have higher computational complexity, and may lose certain effects when the motion amplitude is large. Moreover, the imaging platform and observed target are in a state of relative motion during imaging of orbital spacecraft, which leads to changes in the depth of field. Excessive reliance on registration algorithms may cause some pixels to drift, which is not conducive to the retention of multi-view geometric information.

2.2. Multi-Exposure Image Fusion Methods in Dynamic Scenes Based on Deep Learning

In recent years, the utilization of deep learning techniques to address the ghosting issue and information recovery in multi-exposure HDR imaging has drawn extensive attention from researchers and achieved significant success. Kalantari et al. [13] were the first to employ a convolutional neural network as the learning model. They initially utilized optical flow to align the input images, then synthesized the HDR image. However, the aligned images may undergo distortion, leading to the appearance of artifacts. Wu et al. [14] designed an end-to-end MEF network which can directly perform end-to-end training without using optical flow to align images, helping to avoid the phenomenon of distortion caused by wrong optical flow. At the same time, the network can create details that do not exist in some poorly exposed areas of the LDR image; however, the fused image cannot be displayed directly on the digital device except with tone mapping technology. Yan et al. [15] proposed a network based on a Dilated Residual Dense Block (DRDB) to merge the attention-guided feature maps in LDR images, which avoids the appearance of artifacts; however, it is easy to lose details in areas with moving object occlusion or oversaturation. Xiao et al. [16] proposed a fusion network based on optical flow and image attention. This network estimates the optical flow of LDR images using a motion estimation module, encodes the optical flow as flow features, and adaptively combines the information of LDR images using the flow and correlation features. However, the details in high-brightness and low-darkness areas of the fused image are easily lost. Qu et al. [22] proposed a transformer-based multi-exposure image mixing framework, which can effectively fuse image pairs with extreme exposure differences through self-supervised multi-task learning and an adaptive fusion strategy; however, the robustness of this approach to issues such as motion blur, occlusion, illumination changes, and noise is not high.

There are a number of general image fusion methods based on unsupervised learning that can fuse visible optical and infrared images and that are applicable to multi-focus and multi-exposure images. Deepfuse, proposed by Prabhakar et al. [19], employs an end-to-end trained deep neural network to merge multi-exposure images into an HDR image, emphasizing detail preservation and visual quality enhancement. However, it may suffer from increased computational complexity and potential artifacts in regions with high dynamic range. Densefuse, proposed by Liu et al. [20], can achieve impressive fusion effects in preserving edge information, but has difficulties in maintaining color consistency and is sensitive to noise in the source image. U2fusion, proposed by Zhang et al. [21], can preserve spectral and spatial details when fusing high-resolution images, but performs poorly when dealing with low-resolution images and has high computational requirements.

The deep learning-based dynamic MEF methods mentioned above can effectively reduce motion artifacts in reconstructed images, and provide fusion effects and runtime efficiency that far surpass traditional methods; however, most of these deep learning-based methods require supervised training using ground truth data, and the training sets only

contain natural scenes. Due to the difficulty of acquiring data for space-based optical images of spacecraft, there are currently no publicly available datasets suitable for multi-exposure image fusion tasks, making supervised methods difficult to apply. Unsupervised methods often consist of general image fusion frameworks, which typically choose a direct addition fusion strategy and have difficulties handling a large scale of moving pixels. When dealing with complex overlapping areas containing detailed information, this approach can lead to more severe ghosting. Additionally, when considering the needs of subsequent tasks such as attitude estimation in the IOS system, it is necessary to strictly retain the original multi-view geometric relationship of the observed target in the image. In order to overcome the above problems, we propose an unsupervised learning method to solve the multi-exposure image fusion problem in IOS scenes. In addition, we construct a multi-exposure spacecraft image dataset that conforms to the orbital rendezvous observation. Our approach enables the creation of high-quality HDR images without altering the representation of the observation targets (such as the angle of a sailboard), leading to better processing effects in subsequent spacecraft detection tasks.

3. Methodology

3.1. Overall Framework

Our multi-exposure fusion method for space-based optical images of the observation target is aimed at achieving spatial alignment of dynamic spacecraft images at the feature level with an alignment module. It can reduce the ghosting generated in the fusion process of complex overlapping areas in spacecraft images. The proposed network is a fully convolutional neural network. Its structure shown in Figure 1. From a functional perspective, this network consists of three parts: a feature extraction module, a feature deformable alignment module, and a feature fusion and reconstruction module.

The feature extraction module is used to transform spacecraft images from the image domain to the corresponding feature space. The feature deformable alignment module is designed for aligning and registering the extracted feature maps one-by-one to eliminate and weaken the spatial information of the dynamic scene on the fusion effect. Dynamic multi-exposure images of spacecraft are quite different from those of general dynamic scenes. During the alignment process, the effective features of spacecraft images only include the dynamic region of the target. In contrast, general images typically have complex static regions. These complex static regions occupy a significant proportion of the feature map, thereby degrading the alignment effect in dynamic areas. The feature fusion and reconstruction module implements information fusion and reconstruction of the aligned features corresponding to different exposures, then outputs the final HDR image.

Simultaneously, our designed loss function incorporates a feature deformable alignment loss function and a feature point extraction loss function. The feature deformable alignment part of the loss function effectively gauges the alignment effect of the feature maps corresponding to overexposed and underexposed images, providing the network with better shape fidelity. Meanwhile, the feature point extraction part of the loss function provides the HDR image with superior computational processing effects, making it more suitable for subsequent tasks of In-Orbit Systems instead of being confined to achieving good visual effects.

For a sequence of dynamic LDR spacecraft images in order $\mathcal{I} = \left\{ I_1, I_2, I_3 \right\}$, the fusion algorithm constructed by our method aims to reconstruct a high-quality HDR image of the observed target that aligns with the reference image I_2 (the image with medium exposure time). The HDR image is expected to contain details with high contrast of three LDR images. Meanwhile, it is not expected to retain ghosting artifacts or change the imaging perspective

of the observed target in the reference image. The generation of the final HDR image H can be represented as follows:

$$H = \Phi(I_1, I_2, I_3; \theta) \tag{1}$$

where Φ represents the fusion network proposed in this paper and θ represents the parameters in Φ.

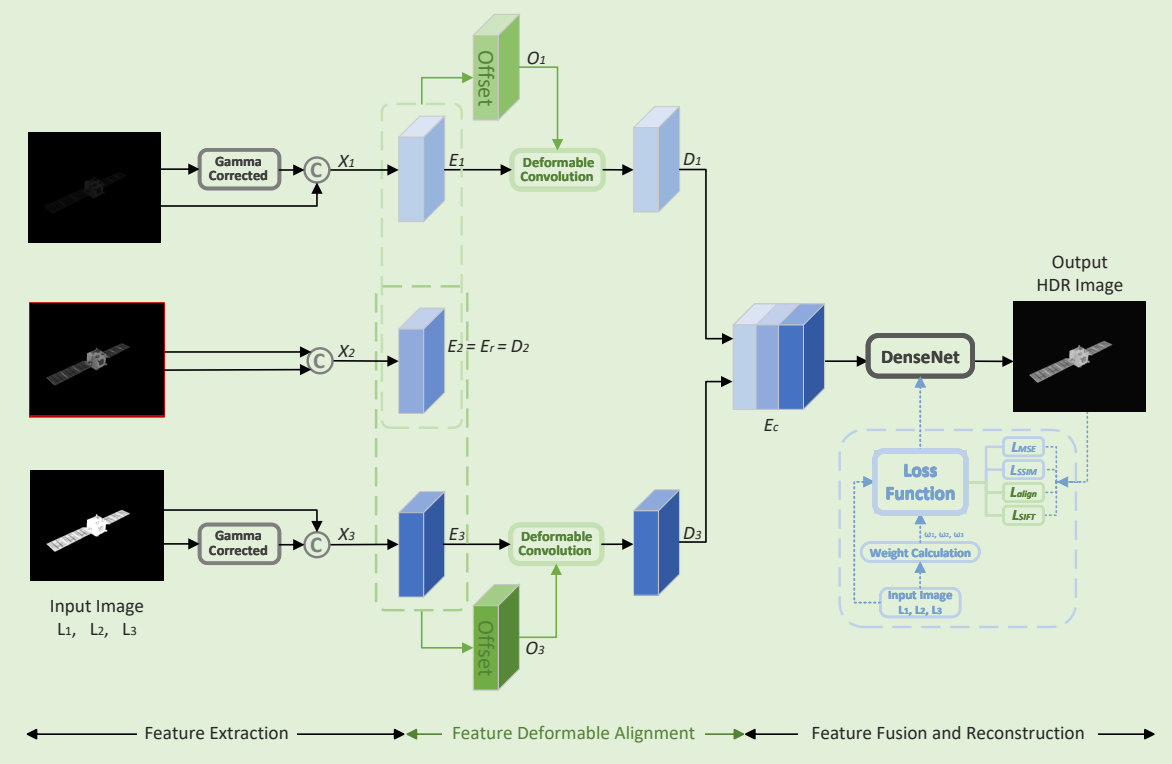


Figure 1. Illustration of the proposed framework. Our framework is composed of three main modules: the feature extraction module, the feature deformable alignment module, and the feature fusion and reconstruction module.

3.1.1. Feature Extraction Module

In general, while LDR images effectively differentiate between exposure times and saturation regions, their ability to assess motion areas is limited. Conversely, images in the linear space of the HDR domain possess a distinct advantage in identifying uncalibrated and dynamic regions due to their simulation of real illumination fields and the establishment of a unified metric for scene representation. Thus, the LDR images are first corrected by a gamma curve to obtain the corresponding sequence $\check{\mathcal{I}} = \left\{\check{I}_1, \check{I}_2, \check{I}_3\right\}$ in the HDR domain:

$$\check{I}_i = \frac{(I_i)^\gamma}{t_i}, i = 1, 2, 3 \tag{2}$$

where t_i represents the exposure time of the ith image and γ represents the gamma correction parameter. Then, each image in the image sequence is input into a six-channel input sequence $\mathcal{X} = \left\{X_1, X_2, X_3 \middle| X_i = [I_i; \check{I}_i], I_i \in \mathcal{I}, \check{I}_i \in \check{\mathcal{I}}, i \in [1, 3]\right\}$ that is connected along the channel dimension, where $\left[\cdot;\cdot\right]$ represents the channel-wise connection submodule.

The stride convolution is introduced for pyramid feature extraction $\mathcal{P}$ from the input sequence $\mathcal{X}$ to transform it into a feature map with 64 channels:

$$E_i = \mathcal{P}(X_i), i = 1, 2, 3. \tag{3}$$

3.1.2. Feature Deformable Alignment Module

We incorporate the deformable convolution [23] into the feature deformable alignment module to accommodate the complex attitude variations of the observed target in spacecraft image sequences. We aim to preserve its multi-view geometric information consistent with the reference image during the process of fusion as much as possible. The deformable convolution can offer the advantage of arbitrary sampling and effectively achieve correction of fundamental geometric transformations such as translation, rotation, scaling, skewing, and perspective distortion. It is suitable for addressing issues related to image sequence mismatches in spatial target monitoring scenarios. Furthermore, its anisotropic bias information facilitates the ability to fit a dense optical flow field, which makes it possible to regulate object deformation dynamics and address nonrigid deformation challenges by modulating both the direction and magnitude of this bias. The feature deformable alignment module computes bias information jointly from both nonreference features E_i and reference features E_r (i.e., E_2). Then, the above information is input into the deformable convolution module to derive spatially aligned features:

$$\begin{aligned} A_i &= dc_i(E_i(p), o_i([E_i(p); E_r(p)])) \\ &= dc_i(E_i(p), O_i(p)) \\ &= dc_i(E_i(p), [\Delta p_j; \Delta m_j]) \\ &= \sum_{j \in \Omega} \omega_j \cdot E_i(p + p_j + \Delta p_j) \cdot \Delta m_j, i = 1, 3 \end{aligned} \tag{4}$$

where p is the position in the feature map, ω_j is the weight of the convolution kernel Ω, and Δp_j and Δm_j respectively represent the learnable bias and modulation strength of the convolution kernel Ω at the jth position.

3.1.3. Feature Fusion and Reconstruction Module

We propose a feature fusion and reconstruction module that combines the aligned feature maps into DenseNet as a single input item for feature fusion and HDR image reconstruction. The feature maps corresponding to each exposure length are obtained through the second module. Regarding the dynamic information of this fixed exposure setting lens, the texture present in adjacent frames is equally rich and significant. Considering that the typical exposure gear is 3, a connected operation is used as the feature merging strategy to obtain the merged feature map $E_c = [A_1; E_r; A_3]$. Then, we send it to DenseNet for fusion and reconstruction to obtain the final HDR image:

$$H = DenseNet(E_c). \tag{5}$$

3.2. Loss

The loss function in this method consists of four parts:

$$\mathcal{L} = \alpha \mathcal{L}_{mse}(\theta, D) + \beta \mathcal{L}_{SSIM}(\theta, D) + \lambda \mathcal{L}_{align}(\theta, D) + \mu \mathcal{L}_{SIFT}(\theta, D) \tag{6}$$

where θ represents the parameters in the network, D represents the training dataset, and α, β, λ, and μ are balanced parameters greater than 0.

3.2.1. Similarity Constraint Loss Function

Because of the lack of ground truth, it is hard to evaluate the quality of our fusion result. However, the Structural Similarity Index Metric (SSIM) can focus on the changes in contrast and structure, while the Mean Squared Error (MSE) can focus on the constraints of intensity distribution differences. Consequently, we implement similarity constraints between the

HDR image and the input LDR images from two perspectives, namely, structural similarity and intensity distribution:

$$\mathcal{L}_{mse}(\theta, D) = \sum_{i=1,2,3} \omega_i MSE(H, I_i), \tag{7}$$

$$\mathcal{L}_{SSIM}(\theta, D) = \sum_{i=1,2,3} \omega_i (1 - SSIM(H, I_i)). \tag{8}$$

The calculation of ω_i references U2fusion [21], which utilizes the VGG network to extract information on the degree of information extraction and information preservation, then uses the softmax function to normalize it such that $\sum_{i=1,2,3} \omega_i = 1$. The weight calculation module in Figure 1 can be represented as $f(\cdot)$, and the overall process of calculating the weight $\mathcal{W} = \left\{\omega_1, \omega_2, \omega_3\right\}$ can be expressed as

$$\mathcal{W} = f(\mathcal{I}). \tag{9}$$

3.2.2. Alignment Loss Function

To achieve better registration of spacecraft image information and preserve the multi-view geometric information of the images as strictly as possible, we propose a feature alignment loss function. If the corresponding static LDR image sequence $\mathcal{I}^* = \left\{I_1^*, I_2^*, I_3^*\right\}$ of the input dynamic LDR image sequence $\left\{I_1, I_2, I_3\right\}$ is known, this means that the viewing angle of the observed target in L_i^* is consistent with the one in I_2, and the exposure intensity is consistent with that in I_i. Therefore, it is only necessary to limit the output feature $\{A_i\}_{i=1}^3$ of the dynamic LDR image feature $\{E_i\}_{i=1}^3$ after correction by the alignment module to approach the feature $\left\{E_i^*\right\}_{i=1}^3$ extracted from the corresponding static LDR image sequence by the encoder as much as possible, thereby achieving the effect of feature alignment:

$$E_i^* = \mathcal{P}(X_i^*). \tag{10}$$

However, many existing dynamic datasets do not include corresponding static images. Therefore, most methods use the reference image to generate pseudo-static scene image sequences $\mathcal{I}^*$ by exposure conversion to assist the learning and updating of network parameters with a self-supervised approach. In contrast, when generating our MES dataset, we directly created a static dataset. This can ensure consistency in both image information and exposure intensity. Rendering images of the Dawn model are shown as an example in Figure 2. Figure 2a–c presents a series of dynamic LDR image sequences, while Figure 2d–f displays their corresponding static LDR image sequences.

Therefore, the loss of the alignment module can be constructed as follows:

$$\mathcal{L}_{align}(\theta, D) = \sum_{i=1,3} (\eta \cdot \|A_i - E_i^*\|_1 + (1 - SSIM(A_i, E_i^*)) \tag{11}$$

where η is the balance parameter. In the alignment loss, $\left\|A_i - E_i^*\right\|_1$ is the fidelity term used to ensure that the spatial information of the feature is as close as possible, while $SSIM(A_i, E_i^*)$ makes the features as consistent as possible in terms of regional and structural characteristics.

3.2.3. Feature Point Extraction Loss Function

Furthermore, we introduce a feature point extraction loss function aimed at optimizing the computer processing effect of the output HDR image for subsequent IOS tasks such as attitude estimation. This loss function is primarily constructed with the SIFT (Scale-Invariant Feature Transform) operator. The SIFT operator is an algorithm employed for feature extraction in image processing and computer vision that detects key points across

various scale-spaces while remaining invariant to image rotation, scaling, and brightness variations. We utilize it in each image with different exposures to identify key points that are expected to possess distinctive local features and exhibit insensitivity to changes in illumination:

$$K_{I_i} = SIFT(I_i), i = 1, 2, 3 \tag{12}$$

$$K_H = SIFT(H) \tag{13}$$

$$n_{I_i} = num(K_{I_i}), i = 1, 2, 3 \tag{14}$$

$$n_H = num(K_H) \tag{15}$$

where $SIFT(\cdot)$ represents the SIFT function for extracting image feature points, K_{I_i} represents the set of feature points in the ith LDR image, K_H represents the set of feature points in the HDR image, $num(\cdot)$ is used to calculate the number of feature points in the set, n_{I_i} represents the number of feature points in the ith LDR image, and n_H represents the number of feature points in the HDR image.

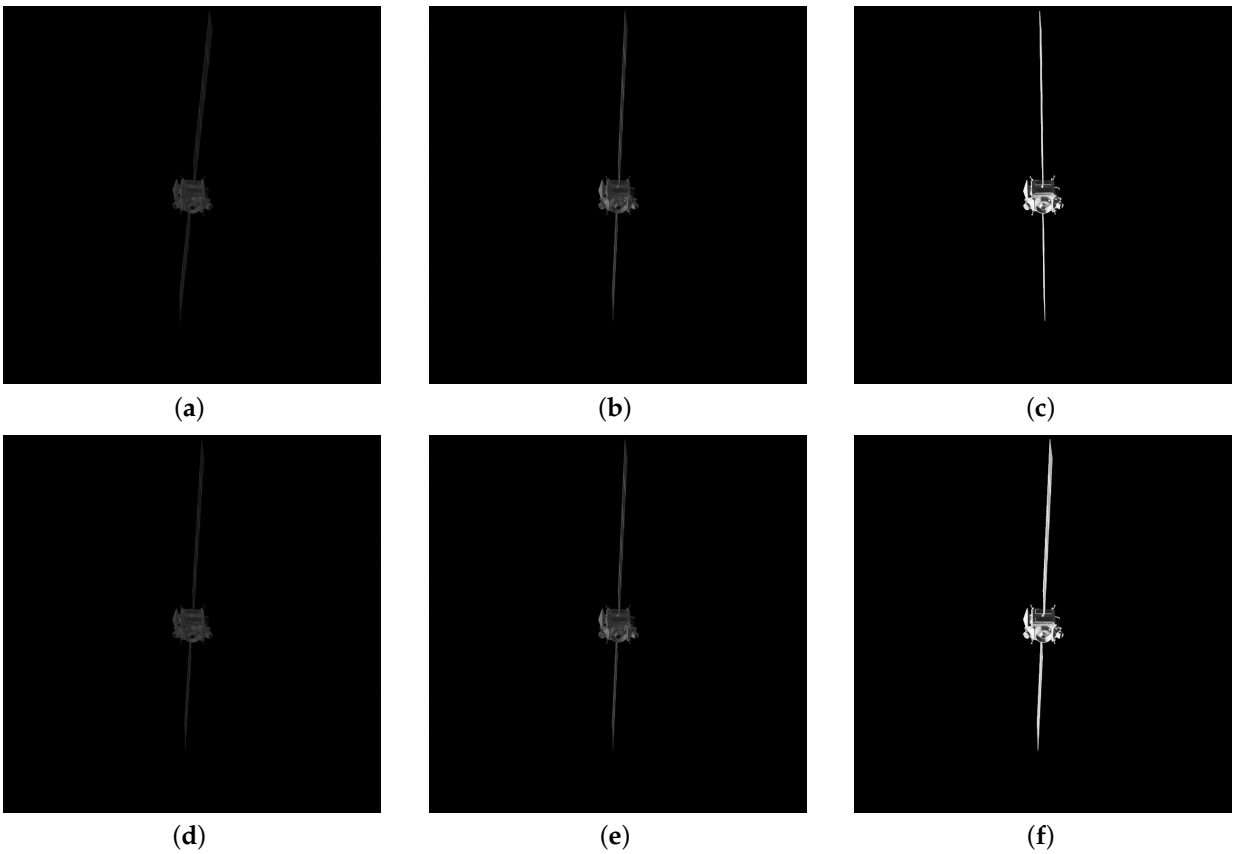


Figure 2. Display of dynamic and static LDR image sequences: (**a**) underexposed image of a dynamic LDR image sequence, (**b**) mid-exposed image of the dynamic LDR image sequence, (**c**) overexposed image of the dynamic LDR image sequence, (**d**) underexposed image of the corresponding static LDR image sequence, (**e**) mid-exposed image of the corresponding static LDR image sequence, and (**f**) overexposed image of the corresponding static LDR image sequence.

To enhance the computer processing effect of the reconstructed HDR image in subsequent space monitoring tasks, the fusion image should incorporate as many unique feature points as possible. Consequently, we take the intersection of the feature point sets from the LDR and HDR images that have undergone redundancy detection and compare

their respective feature point counts. In this way, we develop the following feature point extraction loss function:

$$\mathcal{L}_{SIFT}(\theta, D) = \frac{1}{1 + e^{\frac{num(rd(K_H)) - num(\max\limits_{\{i=1,2,3\}}(rd(K_{I_i})))}{num(rd(\bigcup K_{I_i})) - num(\max\limits_{\{i=1,2,3\}}(rd(K_{I_i})))}}} \tag{16}$$

where $rd(\cdot)$ denotes the detection of repetition in the feature point set, $rd(K_H)$ denotes the set of feature points after repetitive detection of the HDR image, $rd(\bigcup K_{I_i})$ denotes the intersection of all non-repeated feature point sets of the LDR sequence, and $\max\limits_{\{i=1,2,3\}}(rd(K_{I_i}))$ denotes the set with the largest number of feature points retained after repeated detection of each feature point set of the LDR sequence. This can be understood numerically as follows: when the number of non-repeating feature points in the HDR image is equal to the number of non-repeating feature points in the LDR image, $\mathcal{L}_{SIFT}(\theta, D) = 0.5$; when the number of non-repeating feature points in the HDR image is less than the number of all non-repeating feature points in the LDR image, then $\mathcal{L}_{SIFT}(\theta, D) = \frac{1}{1+e}$. That is, the closer the value of the loss function is to 1, the fewer non-repeating feature points there are in the HDR image, while the closer the value of the loss function is to 0, the more non-repeating feature points there are in the HDR image.

4. Dataset

Due to the difficulty of obtaining actual observation images of spacecraft, most image datasets are generated through software rendering or semi-physical simulation, such as the BUAA-SID dataset [24] and the SPEED+ dataset [25]. However, existing simulation datasets do not consider space-based imaging scenarios of spacecraft with the specific photographic technique called exposure bracketing performed during the orbital rendezvous. At the same time, there are changes in the illumination condition of the observed target in different regions of the orbit, as well as differences in the reflectivity of its various components. Therefore, we used Blender to create a multi-exposure image dataset of spacecraft observed during orbit intersection. The simulation steps for generating spacecraft images are shown in Figure 3.

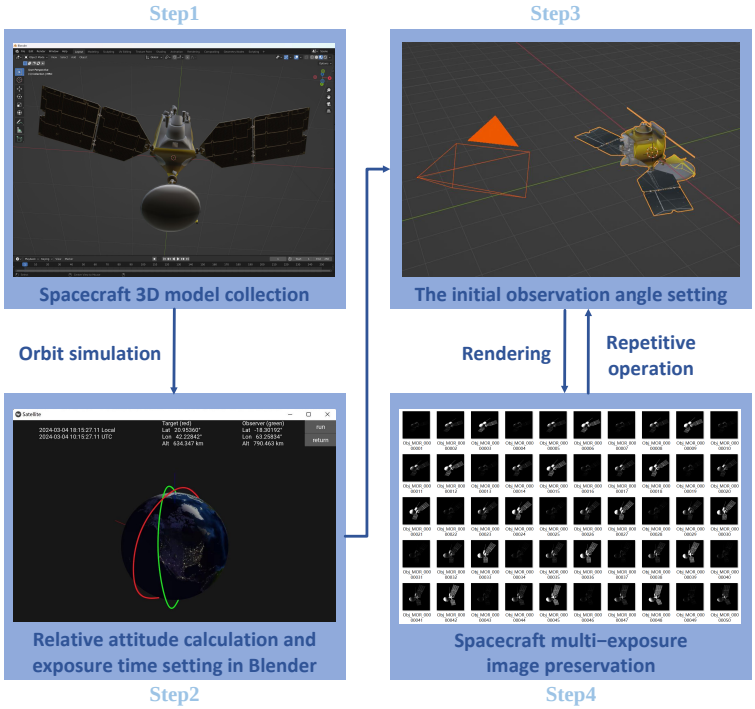


Figure 3. The production process of the MES dataset.

4.1. Spacecraft Model Collection

Our dataset employs twelve spacecraft models obtained from publicly accessible online resources such as NASA's Open Source Satellite 3D Model website. First, we converted all space object 3D models to .obj format in order to import them into Blender, facilitating automated processing with Python scripts. It is worth noting that the materials and refractive indices of different parts of the space objects were different. Exposure bracketing was used when imaging, which can obtain the details of the observed target under different exposure conditions. To restore the actual imaging situation as much as possible, we adjusted the material and refractive index parameters of each model we collected.

4.2. Rendering Process

Blender is a 3D modeling and animation rendering software that can be used to generate spacecraft observation images that are similar to real scenes. To simulate the real environment of space and the actual scenario of orbital rendezvous, we set the scene background to black, the light source to a plane light source, and the camera position to the relative position between the imaging platform and the observation target. We achieved exposure bracketing by setting the imaging time.

For spacecraft and observation platforms with different orbital information, the imaging moment of each frame during the orbital rendezvous is different. When the orbital information is known, it is possible to calculate the relative position of the sun and the observation target at the current moment during the observation period in order to simulate the light intensity in a real space scenario. We obtained images with low, medium, and high exposure times by changing the shutter time setting. The relative position and distance between the imaging platform and the observation target can also be calculated using orbital parameters. The light energy intensity and lens focal length in the simulation software cannot be set the same as those of the sun and the observation platform. To simulate the actual observation effect as much as possible, we scaled the real parameters after calculation for the rendering software simulation parameter settings.

Because the actual observation times of the imaging platform and the observed target are short, the relative attitude variation of the spacecraft in the camera coordinate system during this observation period is minimal. To maximize the richness of our training data, we generated multiple exposure image sequences of a single target with varying poses by adjusting its initial position. This dataset is named the Multi-Exposure Spacecraft Image Dataset (MES dataset), with selected images illustrated in Figure 4. The MES dataset includes multi-exposure image sequences with four initial poses for each of the twelve spacecraft models. There are a total of 48 sequences, each with 120 images; thus, it includes 1920 multi-exposure sets, equating to 5760 individual images. We used images from eight of the spacecraft models for training and used the other four for testing.

To validate the effectiveness of our proposed method under realistic conditions, we conducted evaluations using semi-physical simulation data prepared in the darkroom. The group of semi-physical simulation data consisted of optical image sequences of the "Shenzhou" spacecraft model captured by a Sony camera and scaled down to a certain ratio. The relative attitude between the camera and the model was precisely controlled by the turntable to simulate the observation conditions during orbital rendezvous. A total of 198 images were captured, grouped into 66 multi-exposure image pairings with a size of 2700×2700, as shown in Figure 5. During the test, the dataset was downsampled due to the limitations of the comparison methods.

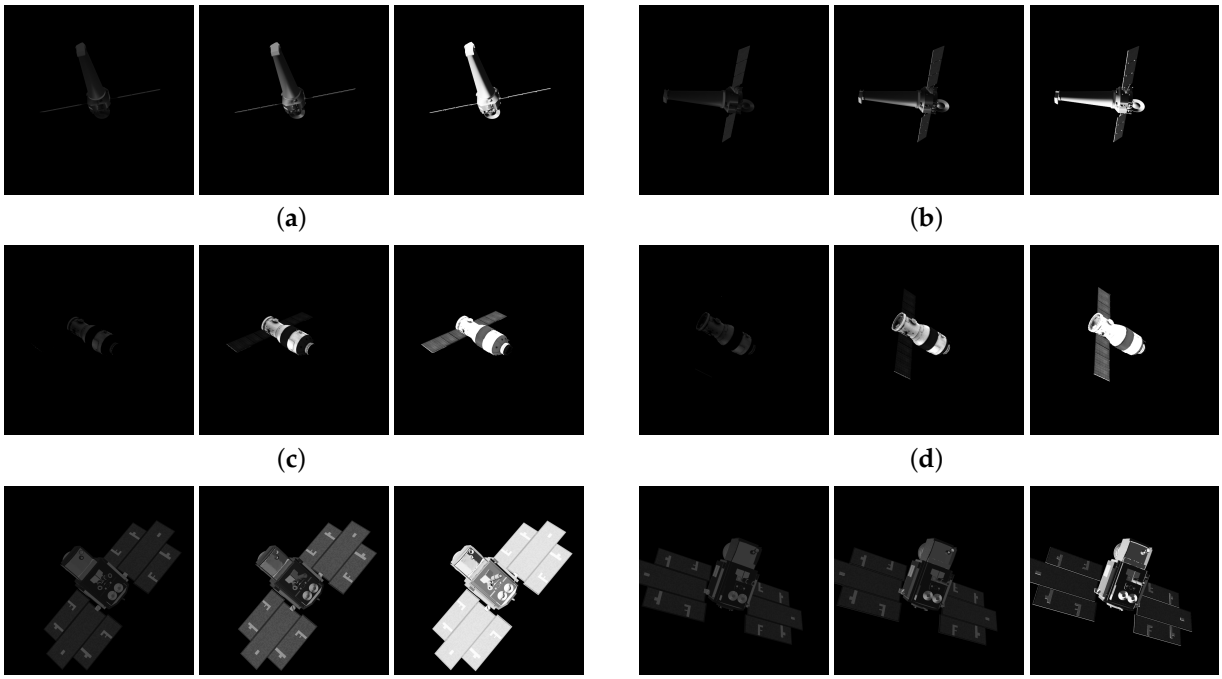


Figure 4. Display of multi-exposure images of different spacecraft in the MES dataset: (**a**,**b**) images of Chandra; (**c**,**d**) images of Dawn; (**e**,**f**) images of Icesat; (**a**,**c**,**e**) respectively represent a set of dynamic multi-exposure images captured at the initial moment of a certain intersection observation for these three distinct spacecraft, while (**b**,**d**,**f**) are the last sets of dynamic multi-exposure images corresponding to the ending moments of the intersection observations of (**a**,**c**,**e**), respectively.

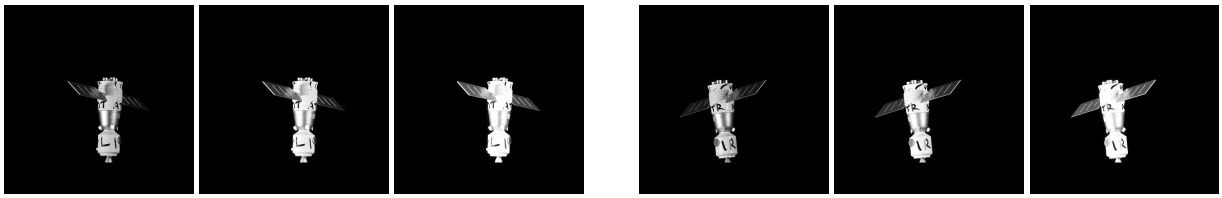


Figure 5. Display of multi-exposure images of Shenzhou spacecraft in the semi-physical simulation dataset: (**a**) represents the set of dynamic multi-exposure images captured at the initial moment of a certain intersection observation, and (**b**) is the last set of dynamic multi-exposure images corresponding to the ending moments of the intersection observations of (**a**).

5. Experiments and Results

5.1. Implementation Details

We trained our network using the MES dataset, including 32 multi-exposure observation sequences of eight spacecraft models, for a total of 1280 multi-exposure image sets with an image block size of 128 × 128. Consequently, during the selection of image blocks, we extracted masks for the spacecraft within high-exposed duration images and randomly selected center point coordinates for these blocks within the mask boundaries until a mean pixel value greater than 80 was achieved for each high-exposure image block. We then extracted a series of 128-length image blocks from the LDR images at these designated center points across all input groups for each round of multi-exposure imaging. This procedure was executed iteratively to ensure both randomness and completeness in feature extraction across all rounds and groups of multi-exposure images. For testing purposes, we utilized sixteen multi-exposure observation sequences involving four spacecraft from the MES dataset, yielding a total of 640 sets comprising 800 × 800 multi-exposure image pairs.

Additionally, we incorporated a set of semi-physical simulation multi-exposure sequences consisting of 66 groups sized at 1350 × 1350 as empirical validation support.

Regarding the specific parameter settings in Equations (6) and (11), we determined the following values through empirical analysis and extensive experimentation: $\alpha = 1$, $\beta = 10$, $\lambda = 1 \times 10^7$, $\mu = 0.2$, $\eta = 1 \times 10^{-10}$. During network training processes, we used the Adam optimizer with default parameters for network training, with an initial learning rate of 1×10^{-4}, a random seed, and 100 training epochs.

All experiments in this paper were conducted on a computer equipped with the following hardware:

- Processor: 12th Gen Intel(R) Core(TM) i9-12900H @ 2.50 GHz (Intel, Santa Clara, CA, USA)
- Graphics Processing Unit: NVIDIA GeForce RTX 3080Ti with 12 GB GDDR6X memory (NVIDIA, Santa Clara, CA, USA)
- Memory: 32 GB

The software environment used in the experiments is as follows:

- Operating System: Ubuntu 22.04
- Deep Learning Framework: PyTorch 1.13.0, with CUDA 11.7 and cuDNN 8.9.6

The comparative methods employed in this study encompassed eight of the most advanced methods for HDR image fusion in dynamic scenes, including five traditional methods (DEM [26], DSIFT [27], DSIFT-EF [28], MEFSIFT [29], SPD-MEF [5]) and three unsupervised learning-based methods (Deepfuse [19], Densefuse [20], U2Fusion [21]). The results from all comparative methods were derived from official open-source code repositories. To ensure a fair comparison, we retrained the unsupervised methods using the same training dataset utilized in this paper.

5.2. Results and Analysis

5.2.1. Objective Evaluation

Because the proposed spacecraft multi-exposure image fusion method is an unsupervised learning one that does not require ground truth images, we only use non-reference quality evaluation metrics for objective evaluation. There are five metrics, including the Correlation Coefficient (CC), Peak Signal-to-Noise Ratio (PSNR), Multi-Exposure Fusion–Structural Similarity index (MEF-SSIM) [30], the number of Extracted Feature Points, and the Attitude Error. It is important to note that the Attitude Error is derived by calculating the mean value of observation sequences for the same target under identical initial attitudes in the camera coordinate system. This metric effectively assesses the degree of change in the multi-view geometric relationships of the observed target within HDR images. In each observation sequence comprising 3*N images, reference images are defined as the N LDR images with medium exposure times, and corresponding N HDR images are matched using feature points to compute relative attitude errors and obtain their average values. To assess whether the multi-view geometric information of the spatial target is effectively preserved, we employ the static image sequence that shares the same exposure as the dynamic image sequence while maintaining the same viewing angle as the reference image to serve as reference inputs for objective quality evaluation metrics. Larger index values of CC, PSNR, MEF-SSIM, and Num-SIFT and smaller index values of AE indicate a better fusion effect.

Table 1 presents the objective quality evaluation numerical results of the proposed method and comparative methods on our MES dataset. Bold values indicate the best results, while underlined values denote the second-best outcomes. In Table 1, it can be seen that the proposed method achieves superior performance in terms of Correlation

Coefficient and Peak Signal-to-Noise Ratio, indicating a strong linear correlation between its output HDR image and the corresponding static image sequence. In addition, it retains more details and structure of the original image while maintaining consistency with the viewing angle of the reference image, with less information loss during the fusion process. Furthermore, our method attains the second-best performance in terms of the MEF-SSIM index, which demonstrates that the reconstructed HDR images exhibit favorable visual quality and effectively retain highlight and shadow details. Finally, our method achieves the best performance in Feature Point Extraction, and its Attitude Error is an order of magnitude smaller than most other methods. These results suggest that the proposed network is well suited for addressing complex relative attitude variations of spacecraft in the camera coordinate system and can be adopted in subsequent tasks of the IOS system. The experimental results indicate that our approach not only retains texture information and contrast without relying on the ground truth, but also offers distinct advantages in preserving the multi-view information of observed targets along with computer processing effects for subsequent technologies such as pose estimation.

Table 1. Mean of the five metrics for the different methods on the MES dataset. Bold values indicate the best results, while underlined values denote the second-best outcomes. ↑ indicates that a larger value of this metric is better and ↓ indicates that a smaller value of this metric is better.

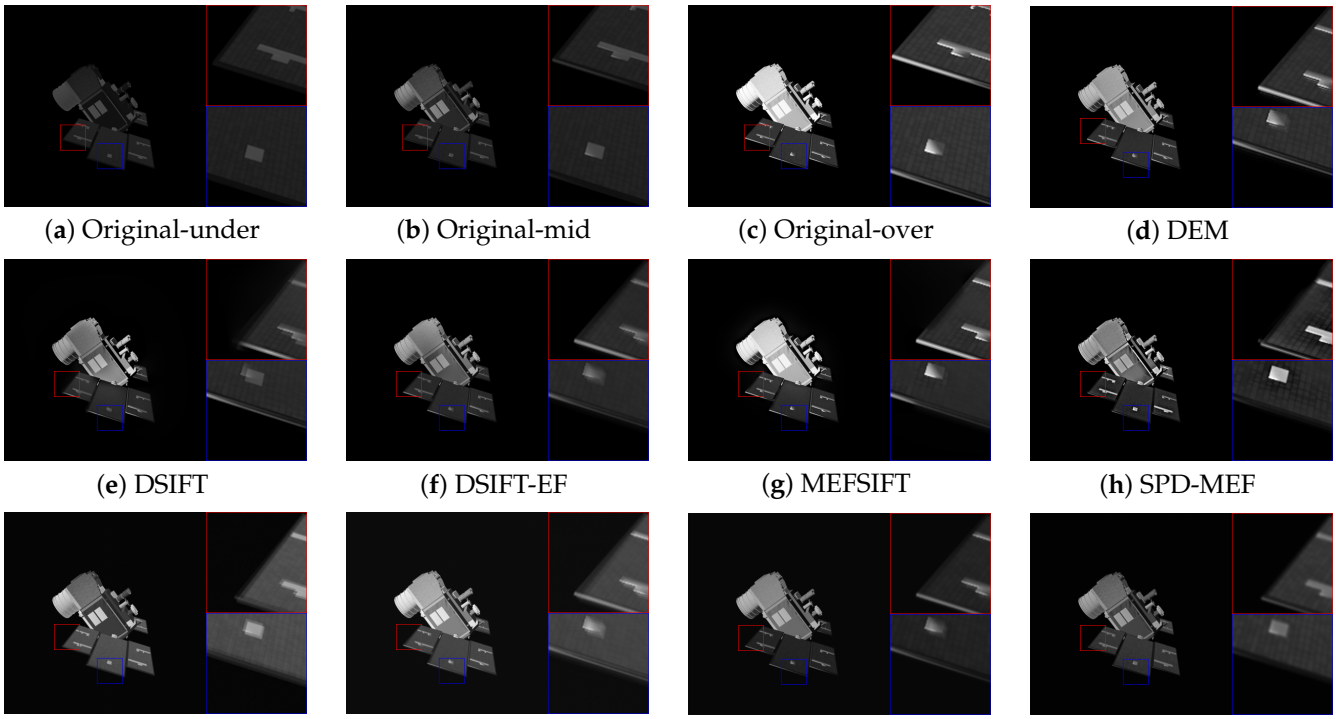
Method	CC ↑	PSNR ↑	MEF-SSIM ↑	Num-SIFT ↑	AE ↓
DEM	0.9035	69.3760	0.9733	38.4667	0.2117
DSIFT	0.8854	68.8791	0.9697	29.6167	0.2154
DSIFT-EF	0.9154	<u>70.1568</u>	0.9746	50.1167	0.1695
MEFSIFT	0.8983	67.8811	0.9723	33.0667	0.4847
SPD-MEF	0.9302	70.0416	**0.9785**	<u>65.9667</u>	<u>0.0315</u>
Deepfuse	<u>0.9445</u>	69.8858	0.9628	50.0833	0.1445
Densefuse	0.9374	68.6223	0.9743	51.0167	0.1689
U2fusion	0.9085	70.0776	0.9671	37.4000	0.2469
Ours	**0.9494**	**70.6305**	<u>0.9752</u>	**80.5167**	**0.0155**

5.2.2. Subjective Results

Figures 6 and 7 illustrate the fusion results and SIFT feature point extraction effects of the proposed method alongside comparative methods on the test set of the MES dataset. As shown in Figure 6, the fusion results of all the compared methods except for SPD-MEF and our method exhibit significant artifacts along the edges and critical regions of the sailboard. However, the fusion result of SPD-MEF has uneven and discontinuous brightness changes at both the sailboard and the main body of the spacecraft. In comparison, our result retains high contrast and texture details. As shown in Figure 7, compared with the other methods, our method can obtain a better feature matching result, especially in the outer sailboard where the relative attitude changes max. Moreover, it is capable of extracting a greater number of feature points while preserving the subjective visual effect, and performs better in the computer processing of subsequent IOS tasks.

Figures 8 and 9 present the fusion results and SIFT feature extraction effects of the proposed method alongside comparative methods on the semi-physical simulation dataset. In Figure 8, it is evident that the fusion results from DEM, DSIFT-EF, Deepfuse, and Densefuse exhibit significant motion blur in the main region, accompanied by noticeable

artifacts at the sailboard. The edge pixels of the sailboard in DSIFT and MEFSIFT display non-smooth variations, while the texture of U2Fusion appears relatively blurred. Although SPD-MEF yields comparatively better fusion results, faint artifacts are still present around the sailboard edge. In contrast to these methods, our method achieves a nice fusion result that is devoid of blur in the main area, with sharp edges on the sailboard that are free from artifacts and retain clear texture details. As illustrated in Figure 9, feature point extraction in the semi-physical simulation images presents greater challenges compared to standard simulation images. In this context, the proposed method demonstrates significantly superior performance when compared to the other methods. These results effectively substantiate the robustness of our method and its strong applicability to semi-physical optical simulations of spacecraft models.

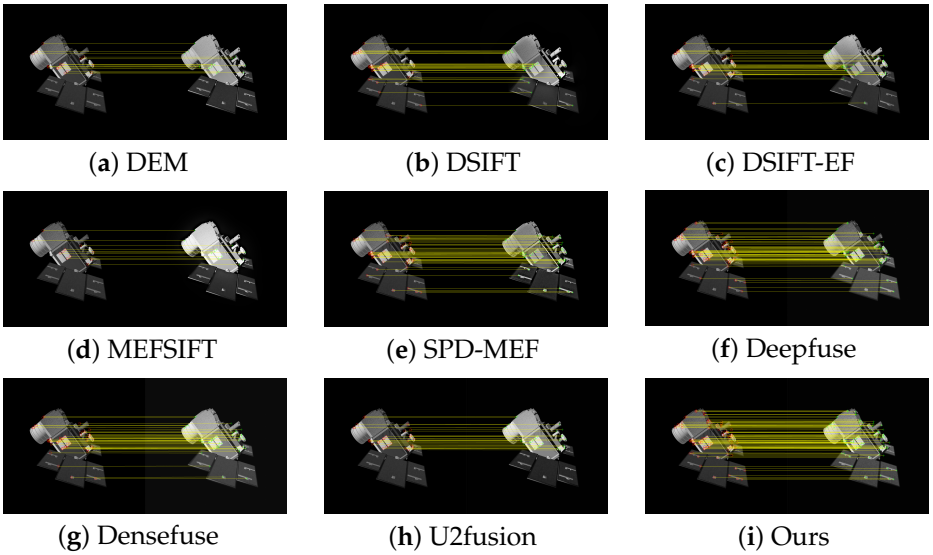


(**a**) Original-under (**b**) Original-mid (**c**) Original-over (**d**) DEM

(**e**) DSIFT (**f**) DSIFT-EF (**g**) MEFSIFT (**h**) SPD-MEF

(**i**) Deepfuse (**j**) Densefuse (**k**) U2fusion (**l**) Ours

Figure 6. Visual comparison of different image fusion methods on the MES dataset.

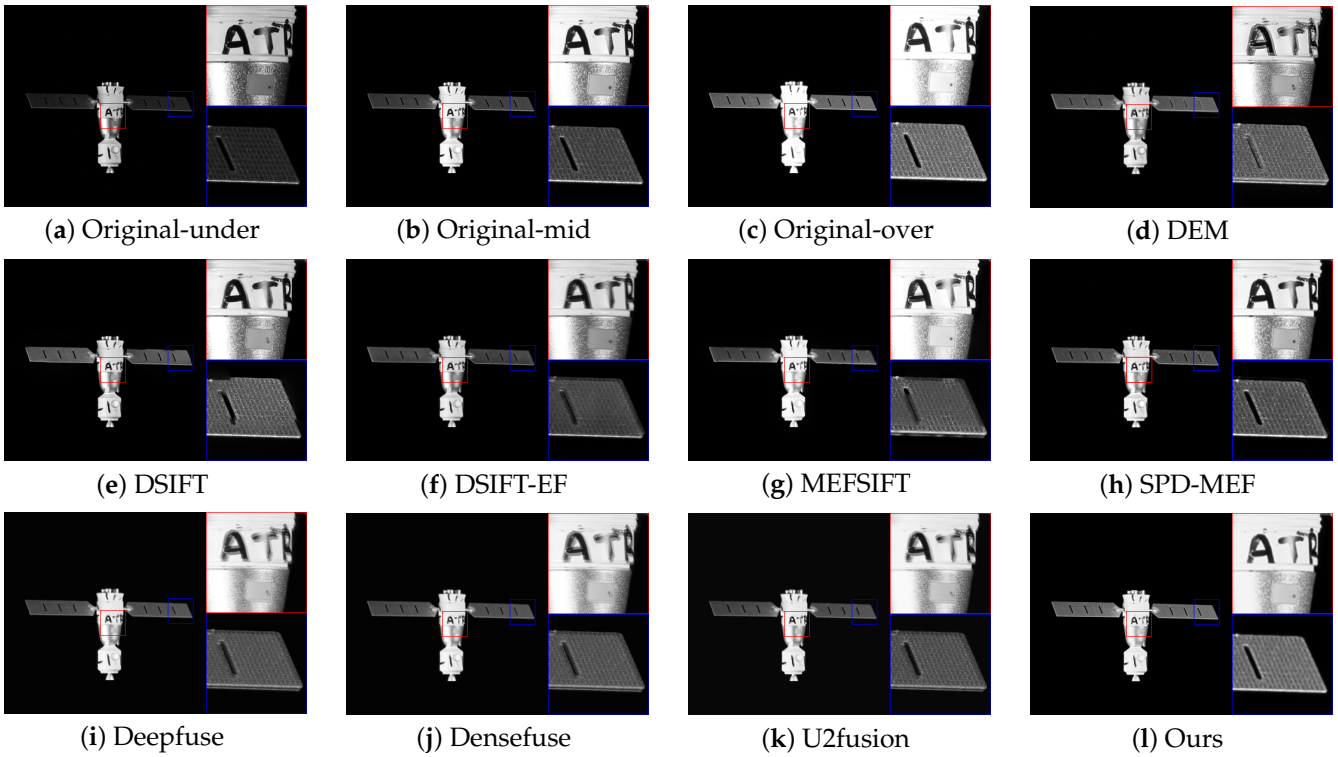


(**a**) DEM (**b**) DSIFT (**c**) DSIFT-EF

(**d**) MEFSIFT (**e**) SPD-MEF (**f**) Deepfuse

(**g**) Densefuse (**h**) U2fusion (**i**) Ours

Figure 7. Visualization of feature point matching on the MES dataset.

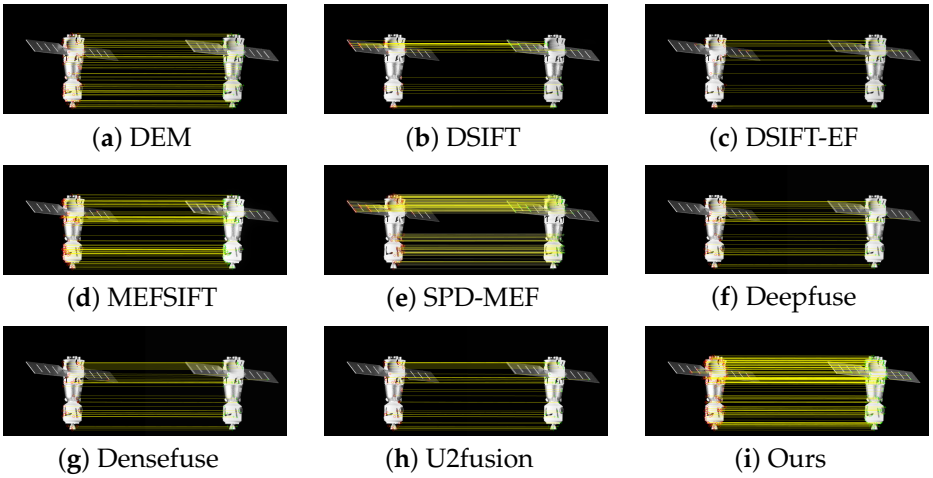


(**a**) Original-under (**b**) Original-mid (**c**) Original-over (**d**) DEM

(**e**) DSIFT (**f**) DSIFT-EF (**g**) MEFSIFT (**h**) SPD-MEF

(**i**) Deepfuse (**j**) Densefuse (**k**) U2fusion (**l**) Ours

Figure 8. Visual comparison of different image fusion methods on the semi-physical simulation dataset.



(**a**) DEM (**b**) DSIFT (**c**) DSIFT-EF

(**d**) MEFSIFT (**e**) SPD-MEF (**f**) Deepfuse

(**g**) Densefuse (**h**) U2fusion (**i**) Ours

Figure 9. Visualization of feature point matching on the semi-physical simulation dataset.

5.3. Ablation Study

5.3.1. Feature Deformable Alignment Module

To assess the optimization effect of the feature deformable alignment module of our network, we conducted separate training sessions for the network with and without the feature deformable alignment module (referred to as FDA in the table) and tested them on our MES dataset. Table 2 illustrates the impact on HDR image fusion performance when employing the feature deformable alignment module, with data in bold indicating the best results. As demonstrated in Table 2, incorporating the feature deformable alignment module enhances all evaluation metrics, doubles the number of extracted feature points, and effectively preserves the representation of observed targets. Additionally, the Attitude Error is reduced by an order of magnitude. These experimental results show that introducing deformable convolution into the feature deformable alignment module and proposing

the alignment loss function allows our network to accommodate the complex relative attitude variations of spacecraft while preserving their multi-view geometric information.

Table 2. Mean of the five metrics on the MES dataset for the proposed network with and without the feature deformable alignment module. Bold values indicate the better result. ↑ indicates that a larger value of this metric is better and ↓ indicates that a smaller value of this metric is better.

Structure	CC ↑	PSNR ↑	MEF-SSIM ↑	Num-SIFT ↑	AE ↓
Without FDA	0.9369	70.5219	0.9555	30.0167	0.1772
With FDA	**0.9494**	**70.6305**	**0.9752**	**80.5167**	**0.0155**

Furthermore, we conducted a visual comparison of the HDR fusion results from the training network with and without the feature deformable alignment module, as illustrated in Figure 10. It is evident that the details in the sailboard region are more pronounced with the incorporation of the feature deformable alignment module and loss function, and the structural edges of the central solar panel appear sharper; additionally, there is a significant enhancement in SIFT feature point extraction performance.

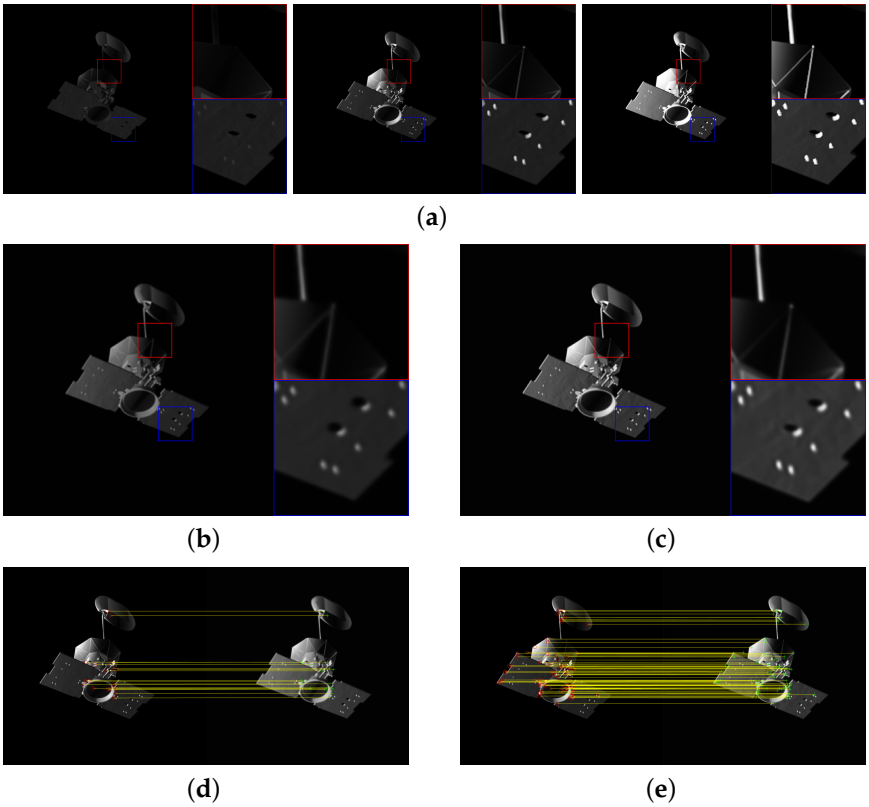


Figure 10. Subjective comparison of our network with and without the feature deformable alignment module: (**a**) the input LDR image set; (**b**,**c**) the output HDR images of our network with and without the feature deformable alignment module; (**d**,**e**) visualization results of feature point matching for our proposed network with and without the feature deformable alignment module.

5.3.2. Feature Point Extraction Loss Function

To assess the optimization effect of the feature point extraction loss function proposed in our method, we conducted separate training sessions for the network with and without $\mathcal{L}_{SIFT}$ and tested the two variants on our MES dataset. Table 3 illustrates the impact on HDR image performance when employing the feature point extraction loss function, with data shown in bold indicating the best results. As shown in Table 3, incorporating the

feature point extraction loss function yields improvements in Correlation Coefficient and Peak Signal-to-Noise Ratio, enhancing the similarity between the fused image and static image sequence. Although there is a slight reduction in MEF-SSIM, this tradeoff allows for greater effectiveness in subsequent computer processing and can better preserve shape information with minimal distortion of observed targets. This results in superior feature point extraction performance and reduces the Attitude Error.

Table 3. Mean of the five metrics on the MES dataset for the proposed network with and without the feature point extraction loss function. Bold values indicate the better result. ↑ indicates that a larger value of this metric is better and ↓ indicates that a smaller value of this metric is better.

Structure	CC ↑	PSNR ↑	MEF-SSIM ↑	Num-SIFT ↑	AE ↓
Without $\mathcal{L}_{SIFT}$	0.9479	69.9931	**0.9756**	74.4000	0.0167
With $\mathcal{L}_{SIFT}$	**0.9494**	**70.6305**	0.9752	**80.5167**	**0.0155**

Furthermore, we conducted a visual comparison of the HDR fusion results with and without the inclusion of the feature extraction loss function in the training network. As illustrated in Figure 11, it is evident that the visual quality of the output HDR images remains nearly identical after incorporating the feature point extraction loss function, even for finer details; in addition, there is an increase in feature matching points on the central solar panel. Consequently, the proposed feature point extraction loss function effectively enhances computer processing performance for reconstructed HDR images while preserving both subjective visual appeal and objective quality for application to subsequent IOS tasks.

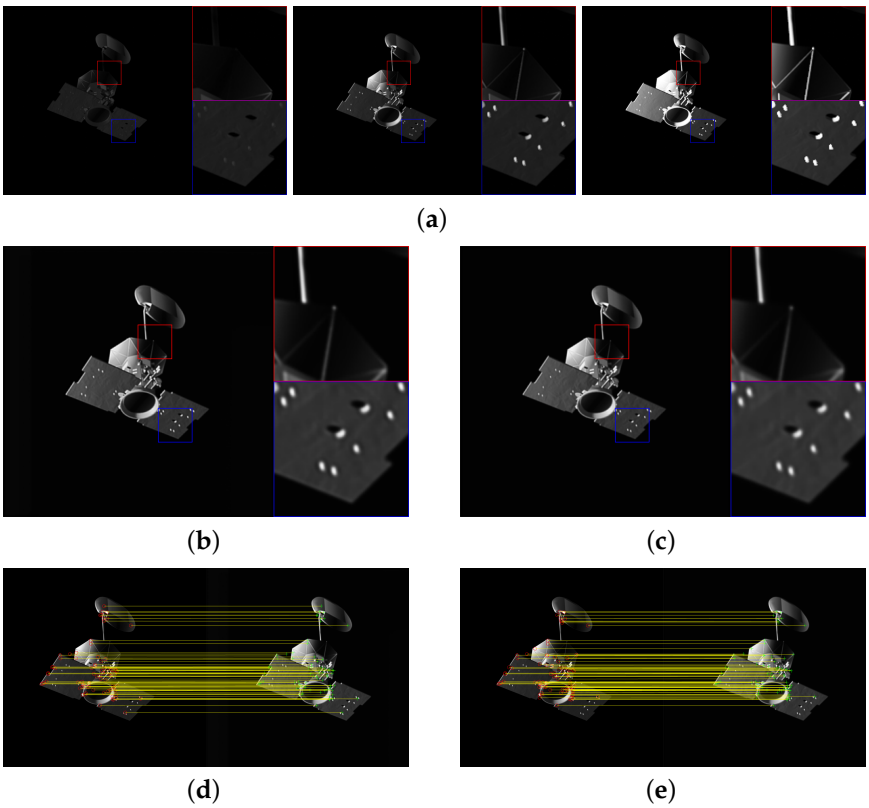


Figure 11. Subjective comparison of our network with and without $\mathcal{L}_{SIFT}$: (**a**) the input LDR image set; (**b**,**c**) the output HDR images of our network with and without $\mathcal{L}_{SIFT}$; (**d**,**e**) the visualization results of feature point matching for our network with and without $\mathcal{L}_{SIFT}$.

5.3.3. Limitations

To explore the limitations of our method, we conducted fusion tests using non-adjacent image frames. The specific settings of the experiment were as follows. First, we took an image with middle exposure duration, denoted as I_k^{mid}, from the kth multi-exposure image set of an observation sequence as the reference image, where $1 < k < 40$. Next, we respectively took i groups of low-exposure-time images I_{k-1}^{low} and high-exposure-time images I_{k+i}^{high} for fusion with I_k^{mid}. In this way, we obtained $i + 1$ comparison groups, with the differences in the number of adjacent frames between comparison groups being $1, 4, 7, \ldots, 3i + 1$ respectively. In order to retain a sufficient number of comparison groups, we set $i = 6$ during the test. As a result, we obtained 896 groups of limitation test data for our subjective and objective analyses.

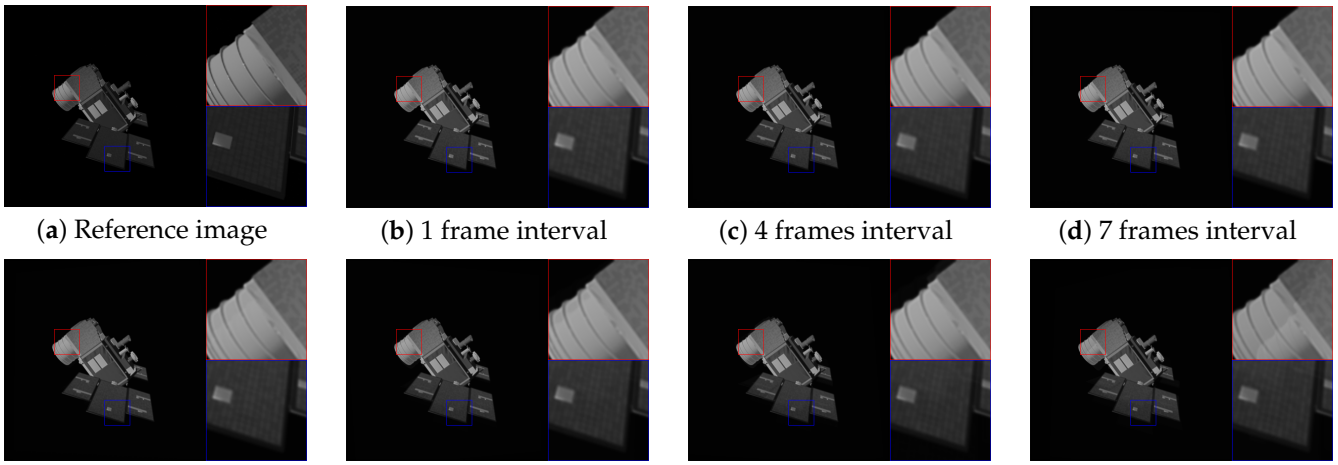
Table 4 presents the objective quality evaluation results corresponding to the fusion results of tthe multi-exposure image groups with different frame differences. As can be seen from Table 4, as the frame difference increases, the attitude error gradually becomes larger. When the frame difference is within four frames, the differences in various indicators are relatively small. Starting from a frame difference of seven frames, the number of feature points gradually decreases. When the frame interval is between thirteen and sixteen frames, all indicators deteriorate significantly.

Table 4. Mean of the five metrics with different numbers of adjacent frames on the MES dataset. Bold values indicate the best results, while underlined values denote the second-best outcomes. ↑ indicates that a larger value of this metric is better and ↓ indicates that a smaller value of this metric is better.

Frame Difference	CC ↑	PSNR ↑	MEF-SSIM ↑	Num-SIFT ↑	AE ↓
1	**0.9494**	73.6342	**0.9751**	<u>80.5445</u>	**0.0166**
4	<u>0.9493</u>	73.6343	0.9742	**80.5781**	<u>0.0174</u>
7	0.9492	73.6341	0.9742	80.5116	0.0177
10	0.9489	<u>73.6345</u>	<u>0.9743</u>	80.4944	0.0197
13	0.9481	**73.6357**	0.9735	80.4688	0.0212
16	0.9473	73.6269	0.9723	73.4949	0.0436
19	0.9374	73.6097	0.9722	75.2447	0.0511

Figure 12 below shows the subjective quality evaluation results corresponding to the fusion results of the multi-exposure image groups with different frame differences. When the frame difference is between one and seven frames, no obvious ghosting appears in the images. When the frame difference is greater than ten frames, the region within the red box in the upper right corner begins to show relatively obvious ghosting. Upon careful observation, it can be seen that the texture color brightness in the blue solar panel area in the lower right corner has changed, and there is a certain inconsistency with the bright and dark areas in (a). This is because the relative attitude change of the solar panel becomes more obvious as the frame difference increases, which can be regarded as an insignificant form of ghosting.

Ultimately, our algorithm exhibits a level of robustness within specific parameters; nonetheless, in instances where there is a pronounced change in relative attitude, we do observe the appearance of some ghosting. Concurrently, the efficiency of feature point extraction and the accuracy of shape fidelity experience a downturn.



(**a**) Reference image (**b**) 1 frame interval (**c**) 4 frames interval (**d**) 7 frames interval

(**e**) 10 frames interval (**f**) 13 frames interval (**g**) 16 frames interval (**h**) 19 frames interval

Figure 12. Visual comparison of our method with frame differences on the MES dataset.

5.4. Results Analysis

According to the objective image quality evaluation results, our method achieves the best performance in terms of the CC, PSNR, Num-SIFT, and AE indicators, and is second-best in terms of MEF-SSIM. When calculating image quality, we used static multi-exposure image sequences as input, meaning that the obtained experimental data can better reflect the shape fidelity of the fusion results. Therefore, the higher the Correlation Coefficient and Peak Signal-to-Noise Ratio, the better our algorithm is at preserving image details and structural information. Notably, it also does well in reducing noise and distortion during the fusion process. This is because the feature deformable alignment module we designed overlaps with the reference image as much as possible during feature alignment, reducing incorrect information mapping.

The Num-SIFT value of our method is 20% higher than that of the second-best method, and at least 50% higher than that of the remaining methods. This is because the feature point extraction loss function we propose can enable HDR images to achieve better computer processing effects. However, this may also result in the loss of some visual perception effects, resulting in our method only obtaining the second-best effect in terms of MEF-SSIM. This conclusion can also be drawn from the results of the ablation experiment described in Section 5.3.2. The most important indicator is the Attitude Error. Our method is far ahead in this indicator, with an error value less than half that of the second-best solution and at least ten times lower than other methods. This result is also attributed to the feature deformable alignment module and feature deformable alignment loss function. Deformable convolutions with arbitrary receptive fields can better adapt to the subtle and complex attitude changes in spacecraft images. At the same time, the setting of the feature deformable alignment loss function is beneficial for feature maps, helping to perceive changes in structural information under the same illumination. These designs enable the fusion results to retain the original multi-view geometric information to the greatest extent. The results of the experiment described in Section 5.3.1 also support this conclusion.

Our method not only performs well in objective evaluation indicators but also has a good subjective visual effect. The fusion results do not contain artifacts at any position in the image, and perform well in feature point extraction and matching results. At the same time, according to the experimental results of the algorithm on the semi-physical simulation dataset, it can be seen that our algorithm has a certain degree of robustness. However, in the limitation experiment, the fusion effect gradually deteriorates as the degree of relative attitude change of the input image group intensifies. This may be because the attitude change between image frames used in training is relatively small. In future work, we will

further optimize our network to allow it to adapt to different degrees of attitude change on the part of space targets. At the same time, we will also consider whether multitask learning could be used to further improve the computer processing effect of HDR images.

6. Conclusions

In this paper, we have presented an unsupervised spacecraft feature deformable alignment network for multi-exposure fusion in dynamic observation scenarios and introduced a multi-exposure dataset for spacecraft. Our method can effectively preserve the original multi-view geometric information of spacecraft images while achieving a pose error less than one-tenth that of most existing methods. In addition, we have designed a loss function based on the SIFT operator which can increase the feature point extraction quantity of HDR images by at least 20%. At the same time, from a subjective visual evaluation, our method does not produce ghosting within a certain range of attitude changes.

Author Contributions: Methodology, Q.X. and X.R.; Software, Q.X.; Data curation, Q.X., H.Y. and C.W.; Writing—original draft, Q.X.; Writing—review & editing, Q.X., X.R., H.Y., L.J., C.W. and Z.W.; Supervision, L.J. and Z.W.; Project administration, L.J. and Z.W.; Funding acquisition, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Xiang, A.; Zhang, L.; Fan, L. Shadow removal of spacecraft images with multi-illumination angles image fusion. *Aerosp. Sci. Technol.* **2023**, *140*, 108453. [CrossRef]
2. Civardi, G.L.; Bechini, M.; Quirino, M.; Colombo, A.; Piccinin, M.; Lavagna, M. Generation of fused visible and thermal-infrared images for uncooperative spacecraft proximity navigation. *Adv. Space Res.* **2024**, *73*, 5501–5520. [CrossRef]
3. Mertens, T.; Kautz, J.; Van Reeth, F. Exposure Fusion: A Simple and Practical Alternative to High Dynamic Range Photography. *Comput. Graph. Forum J. Eur. Assoc. Comput. Graph.* **2009**, *28*, 161–171. [CrossRef]
4. Shen, R.; Cheng, I.; Shi, J.; Basu, A. Generalized Random Walks for Fusion of Multi-Exposure Images. *IEEE Trans. Image Process.* **2011**, *20*, 3634–3646. [CrossRef]
5. Ma, K.; Li, H.; Yong, H.; Wang, Z.; Meng, D.; Zhang, L. Robust Multi-Exposure Image Fusion: A Structural Patch Decomposition Approach. *IEEE Trans. Image Process.* **2017**, *26*, 2519–2532. [CrossRef]
6. Qu, L.; Liu, S.; Wang, M.; Song, Z. Deep unsupervised learning based on color un-referenced loss functions for multi-exposure image fusion. *Inf. Fusion* **2021**, *66*, 18–39.
7. Zhu, J.; Li, J. MSCNNLP: A Multi-scale Convolutional Neural Network and Laplace Pyramid Method for Multi-exposure Image Fusion. *J. Appl. Sci. Eng.* **2020**, *23*, 563–569.
8. Gallo, O.; Gelfandz, N.; Chen, W.C.; Tico, M.; Pulli, K. Artifact-free High Dynamic Range imaging. In Proceedings of the 2009 IEEE International Conference on Computational Photography (ICCP), San Francisco, CA, USA, 16–17 April 2009; pp. 1–7.
9. Wang, Z.; Liu, Q.; Ikenaga, T. Visual Salience and Stack Extension Based Ghost Removal for High-Dynamic-Range Imaging. In Proceedings of the 2017 IEEE International Conference on Image Processing: ICIP 2017, Beijing, China, 17–20 September 2017; pp. 1532–2298.
10. Zimmer, H.; Bruhn, A.; Weickert, J. Freehand HDR Imaging of Moving Scenes with Simultaneous Resolution Enhancement. *Comput. Graph. Forum* **2011**, *30*, 405–414. [CrossRef]
11. Sen, P.; Kalantari, N.K.; Yaesoubi, M.; Darabi, S.; Goldman, D.B.; Shechtman, E. Robust patch-based hdr reconstruction of dynamic scenes. *ACM Trans. Graph.* **2012**, *31*, 203.1–203.11. [CrossRef]
12. Hu, J.; Gallo, O.; Pulli, K.; Sun, X. HDR Deghosting: How to deal with Saturation? In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition: CVPR 2013, Portland, OR, USA, 23–28 June 2013; pp. 747–1489.
13. Kalantari, N.K.; Ramamoorthi, R. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.* **2017**, *36*, 144. [CrossRef]

14. Wu, S.; Xu, J.; Tai, Y.W.; Tang, C.K. Deep High Dynamic Range Imaging with Large Foreground Motions. In Proceedings of the Computer Vision—ECCV 2018: 15th European Conference, Munich, Germany, 8–14 September 2018; Proceedings, p. II; Springer Nature: Cham, Switzerland, 2018; pp. 120–135.
15. Yan, Q.; Gong, D.; Shi, Q.; Hengel, A.V.; Shen, C.; Reid, I.; Zhang, Y. Attention-guided Network for Ghost-free High Dynamic Range Imaging. In Proceedings of the 2019 IEEE/CVP Conference on Computer Vision and Pattern Recognition: CVPR 2019, Long Beach, CA, USA, 15–20 June 2019.
16. Xiao, Y.; Veelaert, P.; Philips, W. Deep HDR Deghosting by Motion-Attention Fusion Network. *Sensors* **2022**, *22*, 7853. [CrossRef] [PubMed]
17. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
18. Niu, Y.; Wu, J.; Liu, W.; Guo, W.; Lau, R.W.H. HDR-GAN: HDR Image Reconstruction From Multi-Exposed LDR Images with Large Motions. *IEEE Trans. Image Process.* **2021**, *30*, 3885–3896. [CrossRef]
19. Ram Prabhakar, K.; Sai Srikar, V.; Venkatesh Babu, R. 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 2017, Venice, Italy, 22–29 October 2017; pp. 4443–5179.
20. Li, H.; Wu, X.J. DenseFuse: A Fusion Approach to Infrared and Visible Images. *IEEE Trans. Image Process.* **2019**, *28*, 2614–2623. [CrossRef] [PubMed]
21. 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] [PubMed]
22. Qu, L.; Liu, S.; Wang, M.; Song, Z. TransMEF: A Transformer-Based Multi-Exposure Image Fusion Framework Using Self-Supervised Multi-Task Learning. In Proceedings of the 36th AAAI Conference on Artificial Intelligence: AAAI-22, Online, 22 February–1 March 2022.
23. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision: ICCV 2017, Venice, Italy, 22–29 October 2017; pp. 745–1490.
24. Zhang, H.; Liu, Z.; Jiang, Z.; An, M.; Zhao, D. BUAA-SID1.0 Space Object Image Dataset. *Spacecr. Recovery Remote Sens.* **2010**, *31*, 65–71.
25. Park, T.H.; Märtens, M.; Lecuyer, G.; Izzo, D.; D'Amico, S. SPEED+: Next-Generation Dataset for Spacecraft Pose Estimation across Domain Gap. In Proceedings of the 2022 IEEE Aerospace Conference: AERO 2022, Big Sky, MT, USA, 5–12 March 2022.
26. Wang, Q.; Chen, W.; Wu, X.; Li, Z. Detail-Enhanced Multi-Scale Exposure Fusion in YUV Color Space. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 2418–2429. [CrossRef]
27. Liu, Y.; Wang, Z. Dense SIFT for ghost-free multi-exposure fusion. *J. Vis. Commun. Image Represent.* **2015**, *31*, 208–224. [CrossRef]
28. Hayat, N.; Imran, M. Ghost-free multi exposure image fusion technique using dense SIFT descriptor and guided filter. *J. Vis. Commun. Image Represent.* **2019**, *62*, 295–308. [CrossRef]
29. 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: 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 1737–1741.
30. Ma, K.; Zeng, K.; Wang, Z. Perceptual Quality Assessment for Multi-Exposure Image Fusion. *IEEE Trans. Image Process.* **2015**, *24*, 3345–3356. [CrossRef]