

## Research Article

# Visible and Infrared Image Fusion-Based Image Quality Enhancement with Applications to Space Debris On-Orbit Surveillance

Jiang Tao <sup>1</sup>, Yunfeng Cao <sup>1</sup>, Meng Ding <sup>2</sup>, and Zhouyu Zhang <sup>1</sup>

<sup>1</sup>College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

<sup>2</sup>College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Correspondence should be addressed to Yunfeng Cao; cyfac@nuaa.edu.cn

Received 26 March 2022; Accepted 1 June 2022; Published 8 July 2022

Academic Editor: Giovanni Palmerini

Copyright © 2022 Jiang Tao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The increasing amount of space debris in recent years has greatly threatened space operation. In order to ensure the safety level of spacecraft, space debris perception via on-orbit visual sensors has become a promising solution. However, the perception capability of visual sensors largely depends on illumination, which tends to be insufficient in dark environments. Since the images captured by visible and infrared sensors are highly complementary in dark environments, a convolutional sparse representation-based visible and infrared image fusion algorithm is proposed in this paper to expand the applicability of visual sensors. In particular, the local contrast measure is applied to obtain the refined weight map for fusing the base layers, which is more robust in a dark space environment. The algorithm can settle two significant problems in space debris surveillance, namely, improving the signal-noise ratio in a noise space environment and preserving more detailed information in a dark space environment. A space debris dataset containing registered visible and infrared images has been purposely created and used for algorithm evaluation. Experimental results demonstrate that the proposed method in this paper is effective for enhancing image qualities and can achieve favorable effects compared to other state-of-the-art algorithms.

## 1. Introduction

Space debris are fragments or elements of invalid man-made space object including spacecrafts and satellites, which are typically generated by space objects' self-explosion, collisions, or end of operational lifetime. According to the statistical data gathered by the Orbital Debris Program Office, most of the observed space debris exists in the low Earth orbit (LEO), which is the region of the most frequent human space operation [1]. Since the existing space debris may cause collision with other operational space objects, the surveillance of existing space debris is crucial for ensuring the safety of human space operation.

Generally, the existing space debris surveillance systems can be classified into two categories, namely, the space-based space surveillance system and ground-based surveillance system. The ground-based surveillance system utilizes a large telescope or radar installed on the ground for space

debris detection and recognition. Different from the ground-based surveillance system, the space-based surveillance system utilizes on-board sensing devices to detect space debris. The advantages of the space-based surveillance system compared to the ground-based surveillance system can be concluded as follows. (1) It is not affected by weather and the circadian rhythm. (2) It could avoid the limits of stationary observation sites. (3) It could detect millimeter-sized small objects, while the latter is aimed at the centimeter-sized objects [2].

Therefore, the development of a space-based surveillance system is effective for enhancing the safety level of a space object. For the application of space debris surveillance, the on-board sensing devices generally contain visible sensors, infrared sensors, and radars. Among all the sensing devices, visible sensors have become a promising solution for the following reasons. (1) Visible sensor can significantly increase the autonomous level of space object observation. (2) It

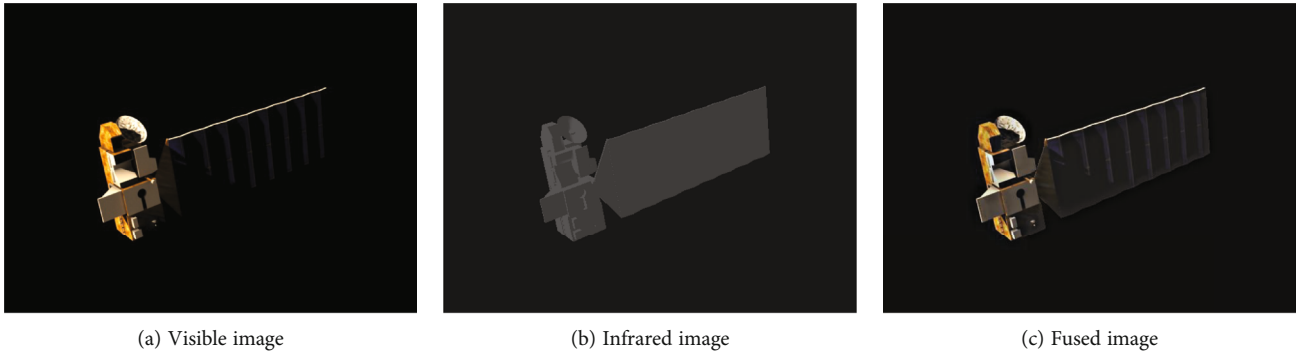


FIGURE 1: Illustration for infrared and visible space debris image fusion.

can provide highly accurate orbital data and detailed texture and edge information of space objects, with precision of up to a millimeter level.

However, the surveillance capability of visible sensors largely depends on the image quality, which can be seriously affected by illumination conditions. Therefore, the effectiveness of visible sensor-based surveillance is largely restricted by illumination. The poor illumination conditions mainly caused by the following two reasons. (1) Space debris's self-obstruction. (2) Chaser's shadow due to the phase angle [3, 4]. Particularly, space debris with insufficient sun irradiation would be observed by visible sensors when they move near the dawn and dusk region. This is defined as weak illumination condition in this paper. As illustrated in Figure 1, the visible parts of the space debris in Figure 1(a) can only reveal platforms apart from the solar panel, while the solar panel is very easy to be ignored in this case. Thus, visible image-based space debris surveillance is insufficient for dealing with the abovementioned challenges. On the other hand, the solar panel of space debris can be captured from the infrared image in Figure 1(b); however, the details within the infrared image including edges and textures are not rich enough compared with the visible image. The fused image in Figure 1(c) takes advantages of infrared and visible images through combining them into a synthetic image.

In order to overcome the abovementioned challenges, infrared sensors are introduced in space perception missions considering the images captured by infrared sensors and visible sensors are highly complementary in weak illumination conditions. Palmerini [5] combined the infrared images and visible images to attain a continuous tracking of the relative position during space proximity operations. Frank et al. [6] used infrared sensors to provide complementary visual information for state estimation during relative navigation. However, these approaches regard the visible images and infrared images as two separate measurements and process them, respectively, which could not make full use of the primitive visual information. In this paper, we proposed a pixel-level infrared and visible image fusion method to enhance image qualities for space debris on-board surveillance. The method could blend the source images sufficiently and retain the primitive information to the utmost extent to expand the applicability of visual sensors in poor illumination conditions of the space perception missions. The algorithm contains three major parts, namely, decomposition,

transformation, and reconstruction. Firstly, both visible and infrared images are decomposed into high-frequency and low-frequency layers. Secondly, the initial weight map is constructed by local contrast measure (LCM) and the refined weight map is further obtained by a guided filter. Then, the low-frequency layers are fused based on the refined weight map. The high-frequency layers are transformed into the sparse domain by utilizing the prelearned convolutional sparse dictionary, and a decision map is generated by evaluating the activity level of transformed sparse coefficients. Then, the high-frequency layers are fused on the basis of the decision map. Finally, the fused image is obtained by synthesizing the high-frequency and low-frequency layers. A dataset containing registered visible and infrared images of space debris is created for algorithm evaluation. Experimental results demonstrate that the algorithm proposed in this paper is effective for enhancing image qualities and can achieve favorable effects compared to other state-of-the-art algorithms.

The rest of the paper is organized as follows. In Section 2, some relevant research is introduced. Section 3 describes the visible and infrared image fusion algorithm in detail. Section 4 presents the evaluation setup. Section 5 shows experimental results for verifying the proposed algorithm. Section 6 concludes the paper.

## 2. Related Works

*2.1. Space-Based Optical Surveillance Techniques for Space Debris.* In this section, the state-of-the-art techniques for space debris surveillance using on-board visible sensors are discussed. In general, according to the size of the space debris image pattern, existing space-based optical surveillance strategies can be classified into two modes: large space debris identification and small space debris detection. The former mode is the application scenario of this paper. During space proximity operations (like space debris removal), the debris is tracked by visible sensors in close range, typically tens of meters to a few centimeters, to identify the object or component's categories (antenna, solar wings, engine nozzle, etc.) and states (attitude, position, spin rate, etc.). Then, different removal strategies are planned based on the space debris categories [7]. Meanwhile, the estimated states are fed into the navigation filter for executing uncooperative rendezvous maneuver during the space debris

TABLE 1: Existing large space debris identification techniques.

Applications	Methods
Category identification	Gang et al. [8], Pan et al. [9], Ding et al. [10], Shi et al. [11], Zhang et al. [12–14], Zhao et al. [15], Yong et al. [16], Tao et al. [17], Dhamani et al. [18]
State estimation	Grompone [19], Sharma et al. [20], Chen et al. [21], Jongh et al. [22], Sonawani et al. [23], Volpe et al. [24]
Object or its component detection	Kanani et al. [25], Petit et al. [26], Wei et al. [27]
D structure construction	Haopeng et al. [28], Zhang et al. [29]

capture. According to different applications, the corresponding methods about large space debris identification are grouped in Table 1, including category identification, state estimation, object or its component detection, and 3D structure construction.

*2.1.1. Category Identification.* Recently, more and more methods concentrate on extracting robust and efficient features, such as handcrafted features and deep learning-based features, to improve the space debris identification performance. Pan et al. [9] proposed invariant moment fusion of the infrared and visible image to recognise the satellite in different sunlight and attitude. Gang et al. [8] combined shaped-based features and appearance-based features to recognise full-viewpoint 3D space objects. Ding et al. [10] characterize the space objects with normalized affine moment invariants (AMI) and the illumination invariant multiscale autoconvolution (MSA) feature descriptor for space object identification with the consideration of viewpoint change, illumination change, and scale change. Shi et al. [11] achieved space object recognition by the bag-of-feature model based on elastic net sparse coding with variation of viewpoints. Zhao et al. [15] applied the sparse coding-based latent semantic feature to recognise the satellite and evaluate it by adding Gaussian white noise. In addition, other machine learning-based space object recognition methods have also been studied. Zhang et al. [13, 14] further used a homeomorphic manifold to represent satellite objects and evaluated the recognition performance in a different lighting phase. Because of the advantages of deep learning in object recognition, convolutional neural network-based space object identification has also been studied recently. Yong et al. [16] and Tao et al. [17] achieved space object recognition using a multilayer convolutional neural network based on LeNet and AlexNet separately. Dhamani et al. [18] selected the MobileNet V1 Single Shot Detector (SSD) architecture for space object detection.

*2.1.2. State Estimation.* Zhang et al. [12] proposed a kernel regression-based method for space object recognition and pose estimation and tested it in different lighting conditions and Gaussian white noise. Grompone [19] detected Harris corner features of the space object for estimating relative linear velocities, angular velocities, and range during docking. Sharma et al. [20] proposed a spacecraft pose network (SPN) method based on the convolutional neural network to estimate the relative position and attitude. Chen et al.

[21] applied High-Resolution Net (HRNet) to obtain a bounding box for estimating the 6DOF pose of a satellite. Sonawani et al. [23] proposed a space object pose estimation method leveraging branch VGG-19 model. Jongh et al. [22] also detected SIFT features of the space debris for pose estimation. Volpe et al. [24] used the KAZE method to detect and describe the features of a tumbling unknown orbital target for determining the relative state. However, the above-mentioned low-level features may ignore the local property or high-level semantic information.

*2.1.3. Object or Its Component Detection.* Wei et al. [27] used shape-based features to detect the spacecraft's components. Kanani et al. [25] and Petit et al. [26] extracted the edge features of debris using bilayer segmentation techniques to distinguish it from the background.

*2.1.4. 3D Structure Construction.* Haopeng et al. [28] proposed a new structure-form motion method to estimate the 3D structure of the space object to avoid the reconstruction error. Zhang et al. [29] extracted scale-invariant feature transform (SIFT) features to recover a 3D structural model of a space object from multiview images.

However, as discussed in the literature, all the aforementioned methods are based on the fact that the illumination condition of space debris is good enough for extracting effective features [30]. Such objects typically have to be large, and the high-resolution image shall be obtained such that they provide abundant details at visible wavelengths, limiting tracking observations to the shadowed regions.

*2.2. Infrared and Visible Image Fusion for Image Quality Enhancement.* Generally, algorithms designed for infrared and visible image fusion include the following three steps: image transformation, image fusion, and image reconstruction. Among the three steps, the method for image transformation is the foundation of the whole algorithm [31]. For this reason, the research of the image fusion algorithm during the past decade mainly focuses on developing a more concise and effective transformation method. The most widely used transformation methods for image fusion are sparse representation (SR), convolutional sparse representation (CSR), and convolutional neural network (CNN).

The application of SR to image fusion has achieved great success in the past few years. However, due to the local representation nature of SR, the drawbacks of the SR-based fusion algorithm are also obvious, which can be concluded

as two manifolds [32, 33]. (1) The context information loss: since SR-based fusion needs to firstly decompose the source image into local patches, the context information within the source image is neglected. It is worth noting that the context information is essential for vision understanding and analysis. (2) The high sensitivity to registration errors: as SR fuses all the image patches, all the image patches need to be accurately registered. However, image registration itself is also a difficult task and the registration error may exist all the time. To overcome this issue, the fusion framework designed on the basis of global representation algorithms is proposed in recent years and the most representative algorithms are the CNN and CSR [31, 34].

The CNN has revealed powerful potential for various computer vision tasks recently. As a supervised learning approach, the framework of the CNN can be classified into two main categories, namely, the regression CNN and classification CNN [35]. Both the regression CNN and classification CNN have been successfully applied to image fusion [36–38]. However, the restriction of CNN-based image fusion may come from the high demand for labeled training samples. CSR is originated from the deconvolutional networks designed for unsupervised image feature analysis. With applications to image fusion, CSR can be treated as an global image transformation approach. The advantages of CSR-based image fusion over SR and the CNN can be concluded as follows [39]. (1) The global modelling capability of CSR makes it free from image decomposition when applied to image fusion. For this reason, the abovementioned deficiencies of SR-based fusion including context information loss and high sensitivity to misregistration caused by local transformation are easy to overcome. (2) The unsupervised learning nature of CSR makes it free from a large amount of labeled ground truth images. Therefore, CSR has revealed great potential for image fusion.

For this reason, the CSR-based method is adopted for image fusion in this paper. Instead of fusing source images directly, both infrared and visible images are decomposed into high-frequency layers and low-frequency layers. The low-frequency layers are fused by the guided filtering-based weighted averaging strategy. The high-frequency layers after decomposition are then transformed into the convolutional sparse domain for image fusion, and the transformed convolutional sparse coefficient maps corresponding to infrared and visible images are fused by activity level assessment. Finally, the fused image is obtained by image reconstruction.

**2.3. The Simulation of Infrared and Visible Space Debris Images.** Due to the dearth of real images of space debris, various methods of image simulation are proposed. Kanani et al. [25] simulated visible images of space debris using Astrium’s in-house tool called Surrender! which is based on rendering functions (rasterization, ray tracing, etc.). Gang et al. [8] created a 3D satellite full-viewpoint visible image dataset by 3ds Max, including a gray image and corresponding binary image. Dhamani et al. [18] created a synthetically generated visible image dataset of the Cygnus vehicle. The model of Cygnus developed in Blender and real images taken from the low Earth orbit are merged into a video game

engine called Unreal Engine 7, to render images with various orientations, lighting conditions, and backgrounds. Volpe et al. [24] used Blender to output the tumbling orbital target with richer textures and more realistic illuminations. Grompone [19] also used open-source software Blender to simulate the visible image of the rendezvous and docking scenario in different lighting conditions, reflections, and background. Sharma et al. [40] used their own camera emulator software of the optical stimulator to render visible images of the Tango spacecraft from the Prototype Research Instruments and Space Mission technology Advancement (PRISMA) mission [41], but the synthesized images all are single-channel grayscale images. Aviles et al. [42] produced both visible and infrared images of the satellite using an ASTOS camera simulator, to study the performance of pose estimation algorithms at a different range of spectrum. Nevertheless, the data is not open source and simulation parameters are not open to public.

As far as we know, there are no public dataset containing registered infrared and visible images of the space debris presently. To assess the quantitative performance of image fusion algorithms, a public space debris dataset with registered infrared and visible images has been created and used for performance evaluation in this paper.

**2.4. Our Contributions.** The contributions of this paper can be divided into three aspects. Firstly, an image quality enhancement framework is proposed for its applications in space debris on-orbit surveillance, which could work in weak illumination condition. Secondly, we propose a colorful visible and infrared image pixel-level fusion method. The method has strong and many desirable properties which are suitable for space debris identification. The local contrast measure-based guided filter is introduced to improve the performance of the original CSR method. Comparing with previous CSR image fusion strategy, our image fusion model can obviously improve the capability of preserving details in faint lighting conditions and enhancing object information in a noise environment. Thirdly, we build a public space debris image dataset named space debris dataset (SDD), which includes registered infrared and visible images. To the author’s best knowledge, it is the first publicly infrared and visible space debris image dataset.

### 3. Infrared and Visible Image Fusion for Space Debris

Figure 2 shows the main process of the proposed infrared and visible image fusion method for space debris surveillance. First, an averaging filter is applied to get the two-scale representations of the source images. Then, the decomposed base layers are fused using a guided filtering-based weighted average method. The guided filter is an edge-preserving smoothing filter which could make full use of the strong correlations between neighborhood pixels for weight optimization [43]. Meanwhile, the decomposed detail layers are fused through convolutional sparse representation models. Finally, the fused image is obtained by combining base layers and detail layers.

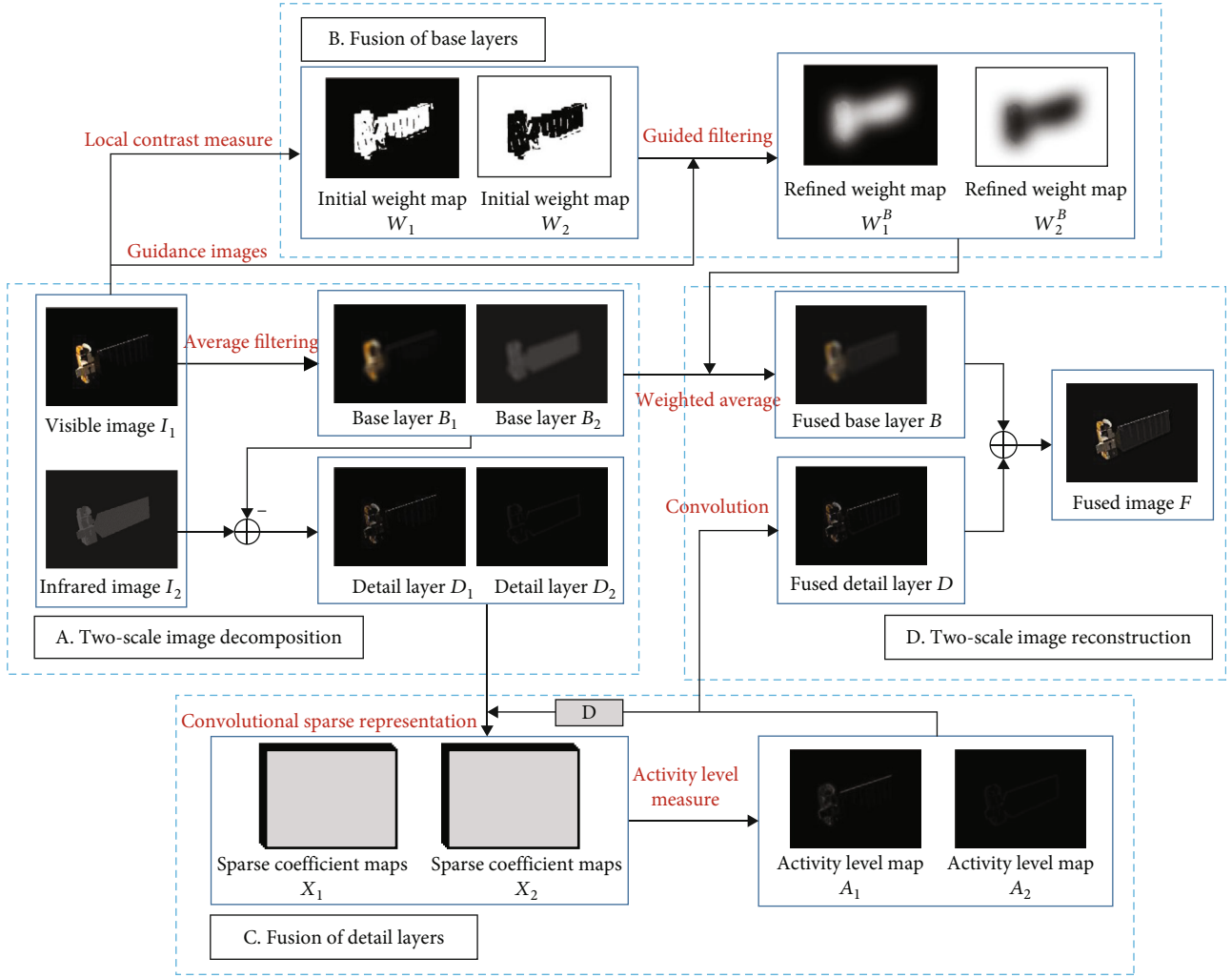


FIGURE 2: Overall pipeline of infrared and visible image fusion model.

**3.1. Two-Scale Image Decomposition.** As shown in Figure 2, both the visible image and the infrared image are decomposed into two-scale representations by an average filter firstly. This step is aimed at splitting the source image into the base layer retaining a low-frequency large-scale image features and detail layer containing small-scale detail features. Suppose that  $I_n$  denotes the  $n$ th source image,  $n \in \{1, \dots, N\}$ .  $A$  represents the averaging filter, and the base layer  $B_n$  of each source image  $I_n$  can be obtained by

$$B_n = I_n A, \quad (1)$$

where the size of the averaging filter is determined by the scale of noise features and desired objects. For example, if the noise features that appeared in the star background are basically less than  $W_N$ , then, the size of averaging filter can be set as  $k \geq 2 \times W_N + 1$  to reduce the noise. On the contrary, if the objects with diameter  $W_O$ , or above, are expected to be reserved, then,  $k \leq 2 \times W_O - 1$  is appropriate. The detail layers  $D_n$  can be obtained by

$$D_n = I_n - B_n. \quad (2)$$

**3.2. Fusion of Base Layers.** An improved guided filtering method is proposed to fuse base layers. The local contrast measure is introduced to enhance the space debris and suppress the background clutters (like stars) that existed in the cosmos. Moreover, this method could preserve color features through applying the guided filtering to the three color channels. As shown in Figure 2, the initial weight map is constructed with local contrast measure firstly. The local contrast value  $C_n$  of the  $n$ th pixel can be calculated by

$$C_n = \min_i L_n \times \frac{L_n}{1/N_c \sum_{j=1}^{N_c} I_j^i}, \quad (3)$$

where  $i$  means one of the nine cells in an image patch obtained by moving a sliding window with size  $w$  on a whole image.  $I_j^i$  represents the gray value of the  $j$ th pixel in the  $i$ th ( $i = 1, 2, \dots, 8$ ) cell except the central cell.  $N_c$  denotes the number of the pixels in the  $i$ th cell.  $L_n$  is the maximum of the gray value within the central cell when the sliding window moves on the  $n$ th pixel. The denominator represents the mean gray value of the  $i$ th cell. The fraction term means the contrast value between the central cell and the

surrounding cell. The obtained local contrast map formed with local contrast values of all the source image pixels is the required initial weight map.

Then, the input image and corresponding guidance image are fed into the guided filter to output the refined weight map. The input image will keep the similar edge characteristics with the guided image by the guided filter, which could also avoid the ringing artifacts in the image decomposition process [43]. Consider the initial weight map  $W_n$  as the input image and color source image  $I_n$  as the guidance image, and the output  $O_i$  of the guided filter can be represented by

$$O_i = \mathbf{a}_k^T \mathbf{I}_i + b_k, \quad \forall i \in \omega_k, \quad (4)$$

where  $i$  means the  $i$ th pixel located in the local window  $\omega_k$  of the guidance image. The local window  $\omega_k$  with size  $(2\alpha + 1) \times (2\alpha + 1)$  is centered at pixel  $k$ . In other words, the premise of linear equation (4) is that the pixel  $i$  must be located in window  $\omega_k$  and the distance between pixel  $i$  and pixel  $k$  should be less than  $\sqrt{2}\alpha$ .  $\alpha$  represents the filter size of the guided filter. The constants  $\mathbf{a}_k$  and  $b_k$  are the linear coefficients of equation (4) when the local window  $\omega_k$  is centered at the  $k$ th pixel. Since the guidance image is a color image,  $\mathbf{I}_i$  is a  $3 \times 1$  vector denoting the pixel values of red, green, and blue channels and  $\mathbf{a}_k$  is also a  $3 \times 1$  vector.  $\mathbf{a}_k$  and  $b_k$  can be calculated by solving the following optimization problem:

$$\operatorname{argmin}_{\mathbf{a}_k, b_k} \sum_{i \in \omega_k} \left( (\mathbf{a}_k^T \mathbf{I}_i + b_k - W_i)^2 + \beta \mathbf{a}_k^2 \right), \quad (5)$$

where  $\beta$  indicates the blur level of the guided filter. It could blur the image details while preserving strong edges of the image.  $W_i$  denotes the pixel value of the  $i$ th pixel of the input image. On the basis of the linear regression theory,  $\mathbf{a}_k$  and  $b_k$  can be given by equations (6) and (7), respectively.

$$\mathbf{a}_k = \left( \sum_{\omega_k} + \beta U \right) \left( \frac{1}{|\omega|} \sum_{i \in \omega_k} \mathbf{I}_i W_i - \mu_k \bar{W}_k \right), \quad (6)$$

$$b_k = \bar{W}_k - \mathbf{a}_k^T \mu_k, \quad (7)$$

where  $\mu_k$  is a  $3 \times 1$  vector denoting a mean gray value of every channel of the guidance image in the local window  $\omega_k$ , and  $\sum_{\omega_k}$  is a  $3 \times 3$  covariance matrix of the guidance image in the local window  $\omega_k$ .  $\bar{W}_k$  refers to the mean gray value of the input image in the local window  $\omega_k$ .  $U$  is a  $3 \times 3$  identity matrix, and  $|\omega|$  represents the number of pixels in the local window  $\omega_k$ . Considering the fact that many filtered output  $O_i$  will be obtained according to different windows  $\omega_k$  covering the  $i$ th pixel, the average

values of  $\mathbf{a}_k$  and  $b_k$  are used to compute the output value of the  $i$ th pixel.

$$\begin{aligned} O_i &= \bar{\mathbf{a}}_i^T \mathbf{I}_i + \bar{b}_i, \\ \bar{\mathbf{a}}_i &= \frac{1}{|\omega|} \sum_{k \in \omega_i} \mathbf{a}_k, \\ \bar{b}_i &= \frac{1}{|\omega|} \sum_{k \in \omega_i} b_k, \end{aligned} \quad (8)$$

where  $\omega_i$  is a local window of the guidance image centered at pixel  $i$  with the same size of  $\omega_k$ . Then, refined weight map  $W_n^B$  composed of filtered output  $O_i$  of every pixel in the guidance image can be obtained. Finally, the fused layer  $B$  can be obtained by weighted averaging as shown in equation (9)

$$B = \sum_{n=1}^N W_n^B B_n. \quad (9)$$

**3.3. Fusion of Detail Layers.** Due to the superior detail preservation ability of the CSR model [33], it is introduced to fuse the detail layers. Provided that  $M$  dictionary filters  $\mathbf{d}_m$  have been learned by the K-SVD method [44], the corresponding sparse coefficient maps  $\mathbf{C}_{n,m}$  of each detail layers  $\mathbf{D}_n$  can be obtained by solving the following optimization problem as shown in equation (10) with the method presented in [45]:

$$\operatorname{argmin}_{\{\mathbf{C}_{n,m}\}} \frac{1}{2} \left\| \sum_{m=1}^M \mathbf{d}_m * \mathbf{C}_{n,m} - \mathbf{D}_n \right\|_2^2 + \lambda \sum_{m=1}^M \|\mathbf{C}_{n,m}\|_1, \quad (10)$$

where  $\mathbf{d}_m$  is a  $d \times d$  matrix,  $d$  is the size of dictionary filters.  $\mathbf{D}_n$  is a  $p \times q \times 3$  tensor, and  $p$  and  $q$  are the height and width of detail layer  $\mathbf{D}_n$ , respectively.  $\mathbf{C}_{n,m}$  has the same dimensions as  $\mathbf{D}_n$ .  $\lambda$  is a preset regularization parameter. By utilizing CSR to obtain the sparse coefficient maps, the activity level measurement fusion strategy based on the  $l_1$  norm is applied to fuse the detail layers. Define  $\mathbf{A}_n(x, y)$  as the activity level map of the  $n$ th source image at the corresponding pixel position  $(x, y)$ , and the initial activity level map  $\mathbf{A}_n(x, y)$  can be obtained by

$$\mathbf{A}_n(x, y) = \|\mathbf{C}_{n,1:M}(x, y)\|_1, \quad (11)$$

where  $\mathbf{C}_{n,1:M}(x, y)$  is an  $M$  dimensional vector denoting the value of  $M$  sparse coefficient maps of the  $n$ th source image at pixel position  $(x, y)$ . Activity level map  $\mathbf{A}_n$  reflects the local energy of sparse coefficient maps of the corresponding source image, which provides quantitative information for assigning the weight to the different source image. It is noteworthy that the fused image would involve unexpected visual artifacts if the source images were misaligned or contain noise. To address these problems, window-based activity (WBA) measurement is employed via fully utilizing the strong correlation among

TABLE 2: Camera properties.

Property	Value
Resolution	752 × 582
Focal length	16.4
Field of view	22° × 16°
F number	0.7
Pixel dimensions	8.6 μm × 8.3 μm
Spectrum range	380 nm–850 nm

adjacent pixels. Then, the final activity level map  $\bar{A}_n(x, y)$  can be obtained by

$$\bar{A}_n(x, y) = \frac{\sum_{a=-r}^r \sum_{b=-r}^r A_n(x+a, y+b)}{(2r+1)^2}, \quad (12)$$

where  $r$  is the size of the window. The larger the  $r$  is, the more details will be eliminated. For the space debris surveillance scenario, stars or cosmic noise with several pixels mainly exists in the background of the space debris image, so a smaller window size might be suitable, and  $r$  is set to 3 in this paper. Next, the appropriate fusion rule is applied for allocating the weight of each source image to the fused image. Averaging and absolute maximum are two most widely used fusion rules; the former could not only retain contrast information but also make the line and edge details more smoother. By contrast, the latter will preserve the most important information of the source images. Thus, the absolute maximum rule is adopted to fuse the detail layers. The fused sparse coefficient maps  $C_{f,m}(x, y)$  can be calculated by equation (13).

$$\begin{aligned} C_{f,m}(x, y) &= C_{n^*,m}(x, y), \\ n^* &= \underset{n}{\operatorname{argmin}} \bar{A}_n(x, y), \end{aligned} \quad (13)$$

where  $n^*$  means the  $n$ th source image corresponding to the max activity level map. At last, the fused detail layer is computed by

$$D = \sum_{m=1}^M \mathbf{d}_m * C_{f,m}. \quad (14)$$

**3.4. Two-Scale Image Reconstruction.** The fused image  $F$  is reconstructed by combining fused base layer  $B$  and fused detail layer  $D$  as shown in equation (15).

$$F = B + D. \quad (15)$$

## 4. Evaluation Setup

**4.1. SDD Dataset.** SDD (<https://github.com/taojianggit/SDD>) is the proposed public available space debris dataset, including 49 pairs of registered visible and infrared images. Each pair of the images is created by using the same camera properties of a close range (CR) camera of the visual based

system (VBS) observing from MANGO spacecraft, which had been demonstrated in Prototype Research Instruments and Space Mission Technology Advancement (PRISMA)-CONTACTLESS deBRIS ACTION (COBRA) experiment [46]. The camera properties are provided in Table 2.

The visible images of space debris are generated by three-dimensional animation software 3Ds-Max Studio. The software uses the raytracer and radiosity technology of Vray engine to render verisimilitude scenes with global illumination. The infrared images are synthesized by Vega Prime software, which utilizes the MOSART Atmospheric Tool to generate time-of-day-dependent atmospheric and material temperature databases for different spectral bands, geographic locations, times of the year, and material lists. It also employs the Texture Material Mapping Tool to create material maps for textures. Although far-infrared bands (8–14) are widely applied for space surveillance due to cost, size, and power consumption of infrared sensors [47], the near-infrared (0.78–3) and mid-infrared (3–8) bands are also considered in SDD to achieve a more comprehensive dataset. In order to generate the weak illumination scene, all the visible images and infrared images are synthesised at the time of twilight, which is set to 05:50 in this work.

As for space debris that resided in the low Earth orbit, the Earth may well occur in the background of space debris images. Therefore, the real Earth images are mixed with synthesised debris images with various spectra, viewpoints, illumination conditions, and ranges to create the dataset. Both real Earth visible and infrared images with different spectra can be obtained from the Himawari-8 geostationary weather satellite (<https://himawari8.nict.go.jp/himawari8-image.htm>). The earth images should be captured from the same illumination viewpoint and altitude as space debris. The altitude of space debris and the surveillance platform are set as 450 km to 500 km in this paper, because many valuable asserts, such as the Chinese Space Station and International Space Station, dwell on there. Besides, the solar glare may appear in the camera of the surveillance platform when space debris is located in an intermediate position. Therefore, it is also considered as one of the backgrounds in the space debris dataset. The visible image of solar glare can be rendered with 3DsMAX Studio. However, according to the characteristic of the solar spectrum, 99 percent of solar radiation concentrates on the wavelength from 0.3 to 3 (<https://www.eia.gov/tools/glossary/index.php?id=Solar%20spectrum>); the infrared images of the sun are ignored in this paper when the discussed spectrum is beyond the range. Besides, stars are not considered to be shown in the background for that they would be swallowed by strong stray light, such as sunlight and Earth and atmosphere radiation (EAR) in the low Earth orbit. The following six typical scenes of space debris surveillance are set in this paper, including ocean, clouds, ocean and clouds, land and clouds, solar glare, and cosmic background.

The SDD includes three categories of space debris: inactive satellite, defunct spacecraft, and rocket body. The inactive satellite model consists of the Tango satellite and Jason-1 satellite. The defunct spacecraft comprises the Tiangong-1 prototype space station. The rocket body is

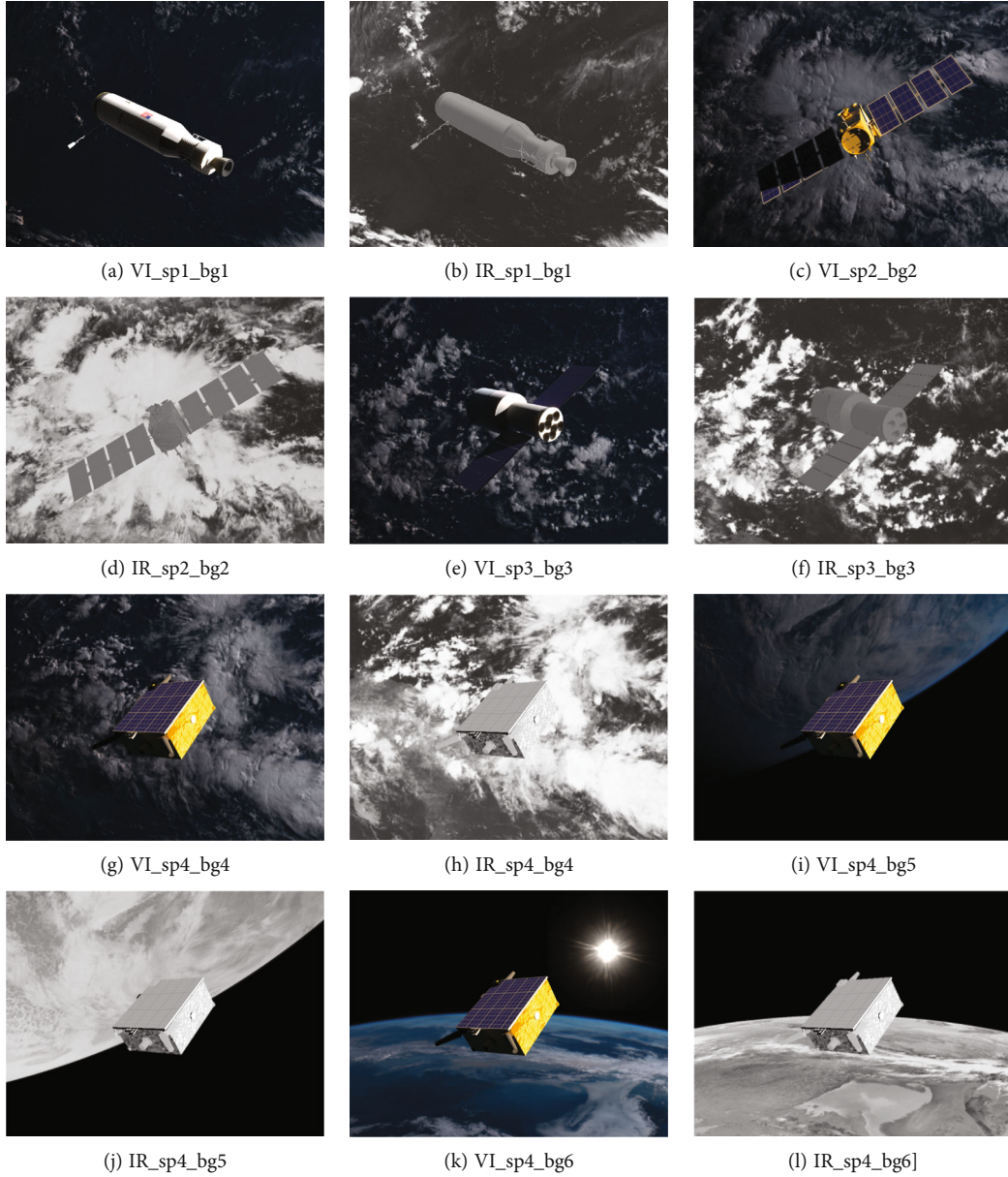


FIGURE 3: Examples of the SDD dataset. “VI” means the visible image, “IR” means the infrared image, “sp” $\{i = 1, 2, 3, 4\}$  means the  $i$ th space debris of SDD, and “bg” $\{i = 1, 2, \dots, 6\}$  means the  $i$ th type of the background of SDD.

composed of the Agena-c rocket upper stage. Thus, 4 kinds of space debris are included in SDD in total. The examples of synthetic visible and infrared images of space debris with different backgrounds are shown in Figure 3.

**4.2. Metrics.** The objective evaluation metrics can be divided into four types in the aspect of image fusion, including information theory-based, image feature-based, image structural similarity-based, and human perception-based metrics [48]. Several typical metrics for each category are selected for evaluating the proposed method comprehensively. Entropy (EN), cross-entropy (CE), mutual information (MI), and peak signal-to-noise ratio (PNSR) of the information theory-based metrics, average gradient (AG), edge intensity (EI), standard deviation (SD), and gradient-based

fusion performance ( $Q^{AB/F}$ ) of the image feature-based metrics, root mean squared error (RMSE) of the image structural similarity-based metric, and Chen-Blum metric ( $Q_{CB}$ ) of the human perception-based metric are adopted. A higher value of the metric shows a better performance of the image fusion method except for CE and RMSE. Apart from the metrics for visible-infrared image fusion, the image quality assessment (IQA) methods are also introduced to sufficiently evaluate the performance of the proposed method. There are three kinds of IQA methods, including full-reference IQA (FR-IQA), reduced-reference IQA (RR-IQA), and no-reference IQA (NR-IQA). Because the reference image of space debris could be synthesised in this work, the FR-IQA metrics are adopted to assess the performance of the proposed method in different datasets. The other



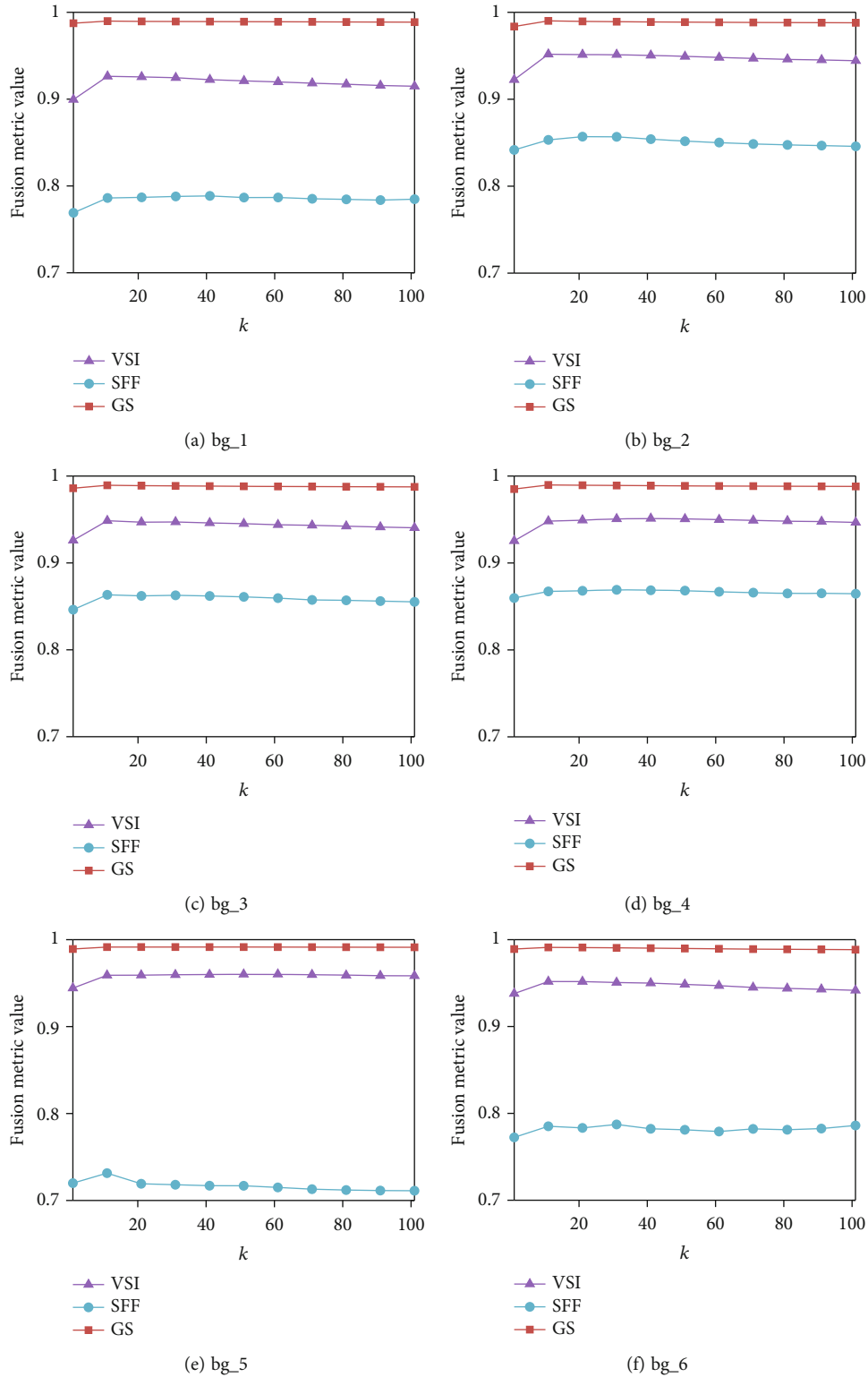


FIGURE 4: Performance of the proposed method with different settings of parameter  $k$  for different backgrounds.

two IQA methods are not applied because they are less credible than FR-IQA. Consequently, the visual saliency-induced index (VSI), sparse feature fidelity (SFF), and gradient similarity (GS) are adopted to assess the proposed method. A large assessment score indicates good performance for these

FR-IQA methods. The details of the abovementioned 13 metrics can be found in [48, 49].

4.3. *Methods for Comparison.* The proposed method is compared with five representative methods, which are the guided

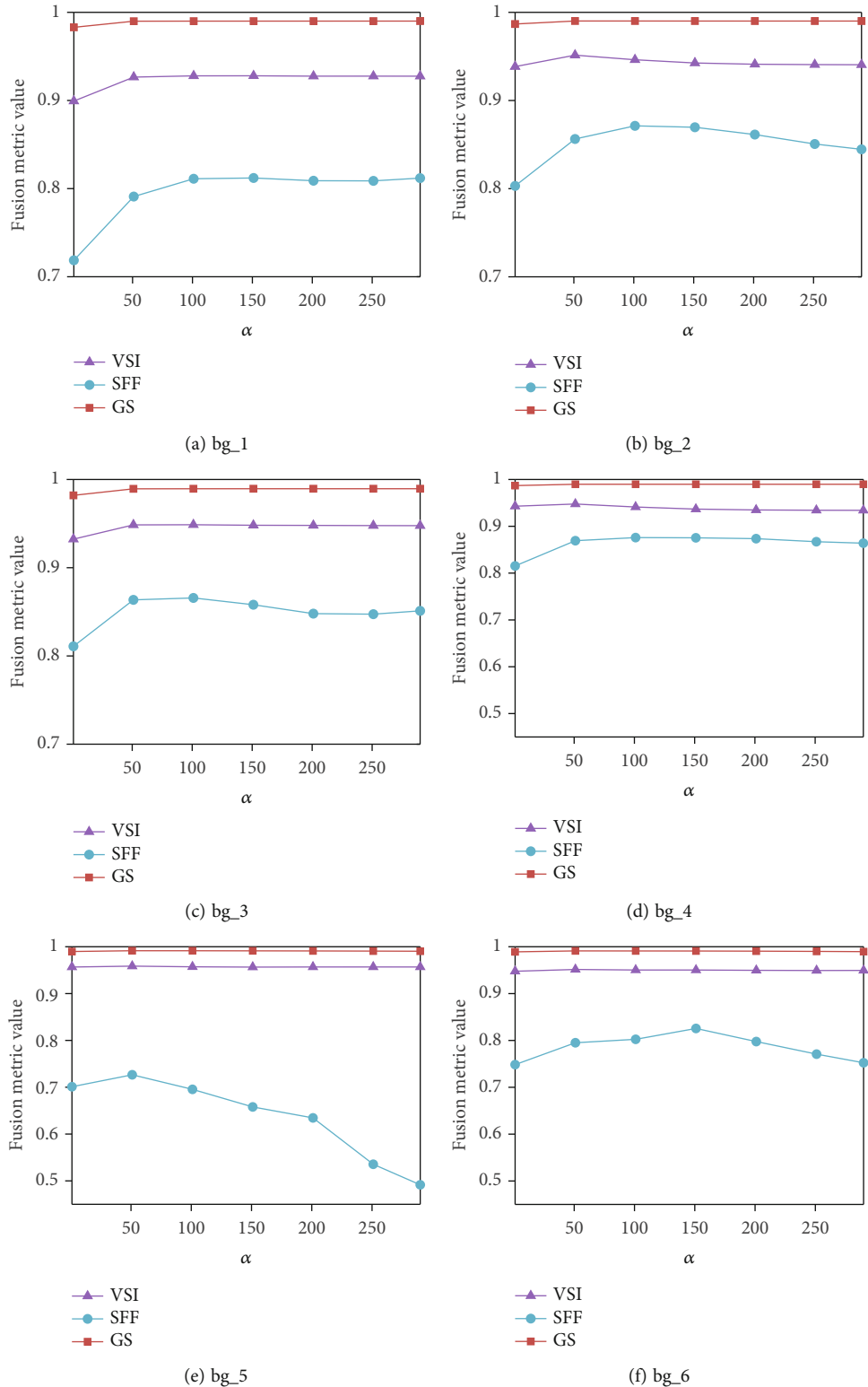


FIGURE 5: Performance of the proposed method with different settings of parameter  $\alpha$  for different backgrounds.

filtering-based fusion (GFF) [43], multiscale GFF (MGFF) [50], convolutional sparse representation (CSR) [33], ResNet [51], and RFN-Nest [34]. Among them, GFF, CSR, and MGFF are three traditional visible-infrared fusion methods. The ResNet and RFN-Nest are representative deep

learning-based visible-infrared fusion methods in the last three years intending to achieve state-of-the-art performance. The implementation of the CSR-based method is available at Liu's website (<https://github.com/yuliu316316/CSR-Fusion>). The implementation of the RFN-Nest method

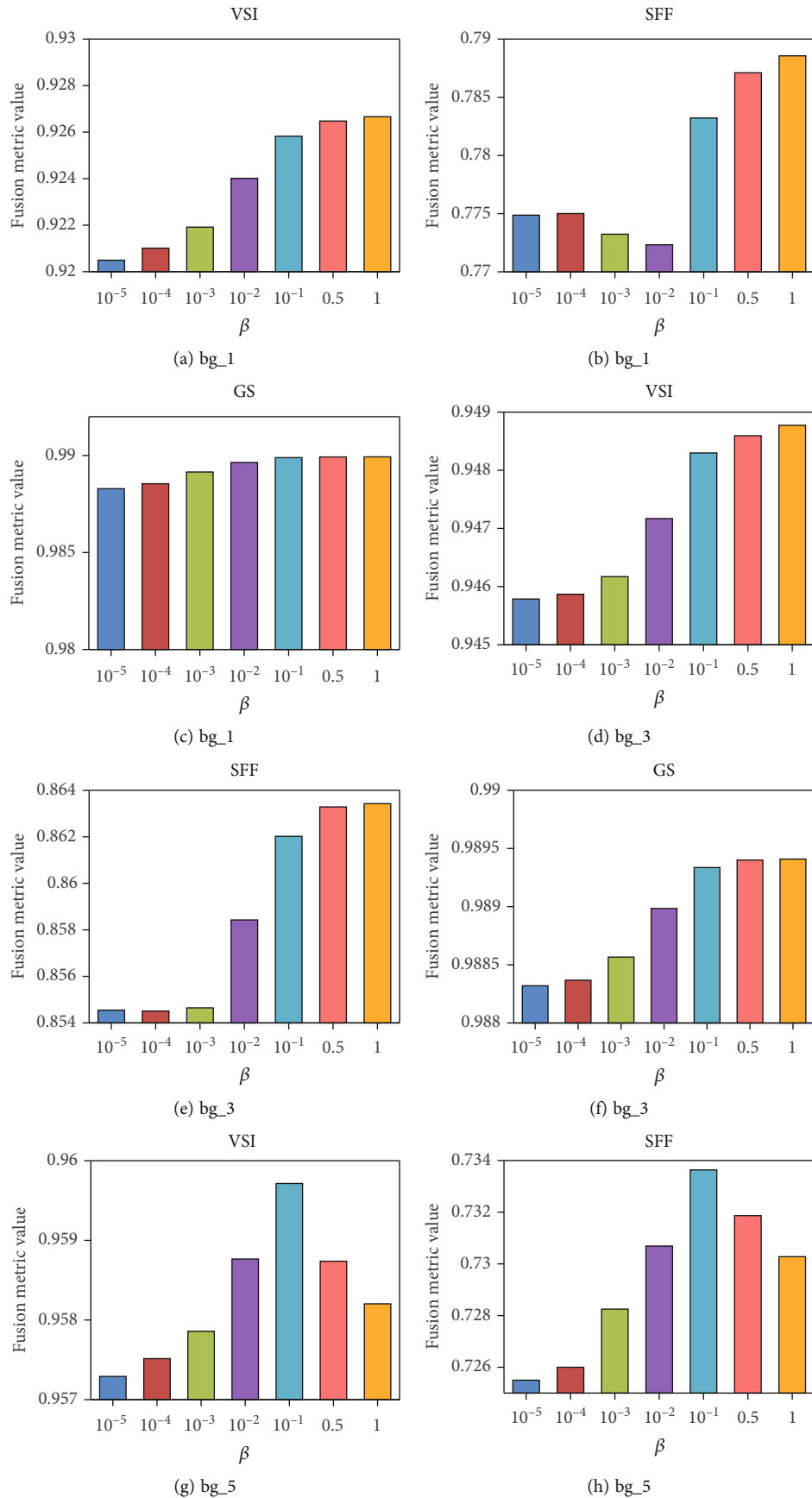


FIGURE 6: Continued.

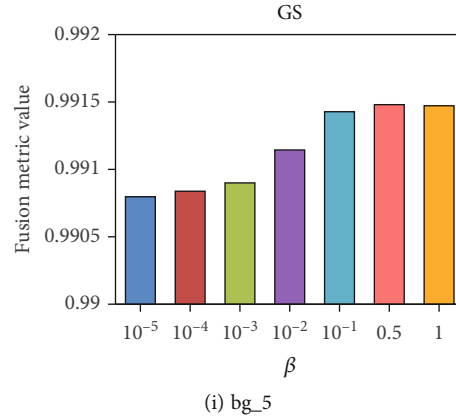


FIGURE 6: Performance of proposed method with different settings of parameter  $\beta$  for different backgrounds.

is available in Li's website (<https://github.com/hli1221/imagefusion-rfn-nest>). All the other methods are implemented by the visible and infrared image fusion benchmark [52]. All the related parameters are set to default values described by the original publications. With regard to the deep learning-based methods, the pretrained models provided by the corresponding authors are adopted.

## 5. Experimental Results and Analysis

This section describes the experimental results on the SDD dataset. Section 5.1 analyses the influence of key parameters. Section 5.2 and Section 5.3 describe the qualitative and quantitative comparison results, respectively. Section 5.4 analyses the comparison results of computational time. All experiments were performed using a computer equipped with an i7-10875H CPU and NVIDIA RTX2060 GPU. Additionally, to verify the robustness of the proposed method, the noise, spectrum, viewpoint, and range are considered to evaluate the performance of different methods in Section 5.5.

**5.1. Parameter Analysis.** In this section, the influence of free parameters of the proposed method are analysed on SDD. The parameters mainly include the size of average filters  $k$ , the size of guided filter  $\alpha$ , and the blur degree  $\beta$  of the guided filter. The evaluation metrics of fusion performance are VSI, SFF, and GSM. To begin with,  $k$  varies from 1 to 101 to test the proposed method. As shown in Figure 4, FR-IQA methods tend to rise at first and then drop. According to the experimental results,  $k$  is set to 11 which is the most appropriate choice to enhance the object and filter the noise for all the six typical backgrounds.

Afterwards,  $\alpha$  and  $\beta$  are analysed in the same way. The fusion metric results are shown in Figures 5 and 6, where  $\alpha$  varies from 1 to 290 and  $\beta$  varies from 0.00001 to 1. As illustrated in Figures 5 and 6, the size of guided filter  $\alpha$  has a similar trend with the size of average filter  $k$ . The blur degree  $\beta$  are analysed on the 1st, 3rd, and 5th kind of background. From the results shown in Figure 6, bigger  $\beta$  is preferred when the base layers are fused. Based on the evaluation metric results, the default parameters are set as

$\alpha = 101$ ,  $\beta = 0.5$ . In this case, a good performance can be obtained for all the six typical scenes.

**5.2. Qualitative Comparison Results.** Figure 7 shows fusion results of different methods on the weak illumination scene. Weak light shone on the surface of the solar panel while strong light shone on the lateral side of space debris in the visible image. The infrared image well captures the front side of space debris, which is not observed in the visible image, but the rich texture information of the solar panel and heat insulating material are mostly observed in the visible image. In this image pair, the desired fusion result is simultaneously preserving the rich textures from the visible image and the details of the front side to enhance the space debris. The fused images of the ResNet method are rather blurry comparing with those of the other method. The RFN-Nest method produces color distortion. Although the GFF method can well preserve the brightness of the solar panel, the contrast of which is not uniform. The CSR method and MGFF method obtain high fusion quality on the whole, but it preserves lower brightness in the front side of space debris than the proposed method.

Based on the abovementioned fusion results, the ResNet method, RFN-Nest method, and GFF method all suffer from drawbacks, like color and contrast distortions, and detail loss. The CSR method and MGFF method obtain competitive performance with the proposed method, but inferior to the proposed method in image brightness. The key factors for the superiority of the proposed method can be summarised as follows: (1) the local contrast measure can enhance the object and suppress the background based on the conspicuousness of object in a local region, and (2) the convolutional sparse representation model can optimize over the entire image which contributes to the good performance of detail preservation.

**5.3. Quantitative Comparison Results.** Table 3 presents the average value of 13 quantitative metrics over 6 source image pairs on SDD in the weak illumination scene. From the results presented in Table 3, the proposed method obtains the best overall performance by giving 4 best values and 2 second best values. The MGFF method obtains the second

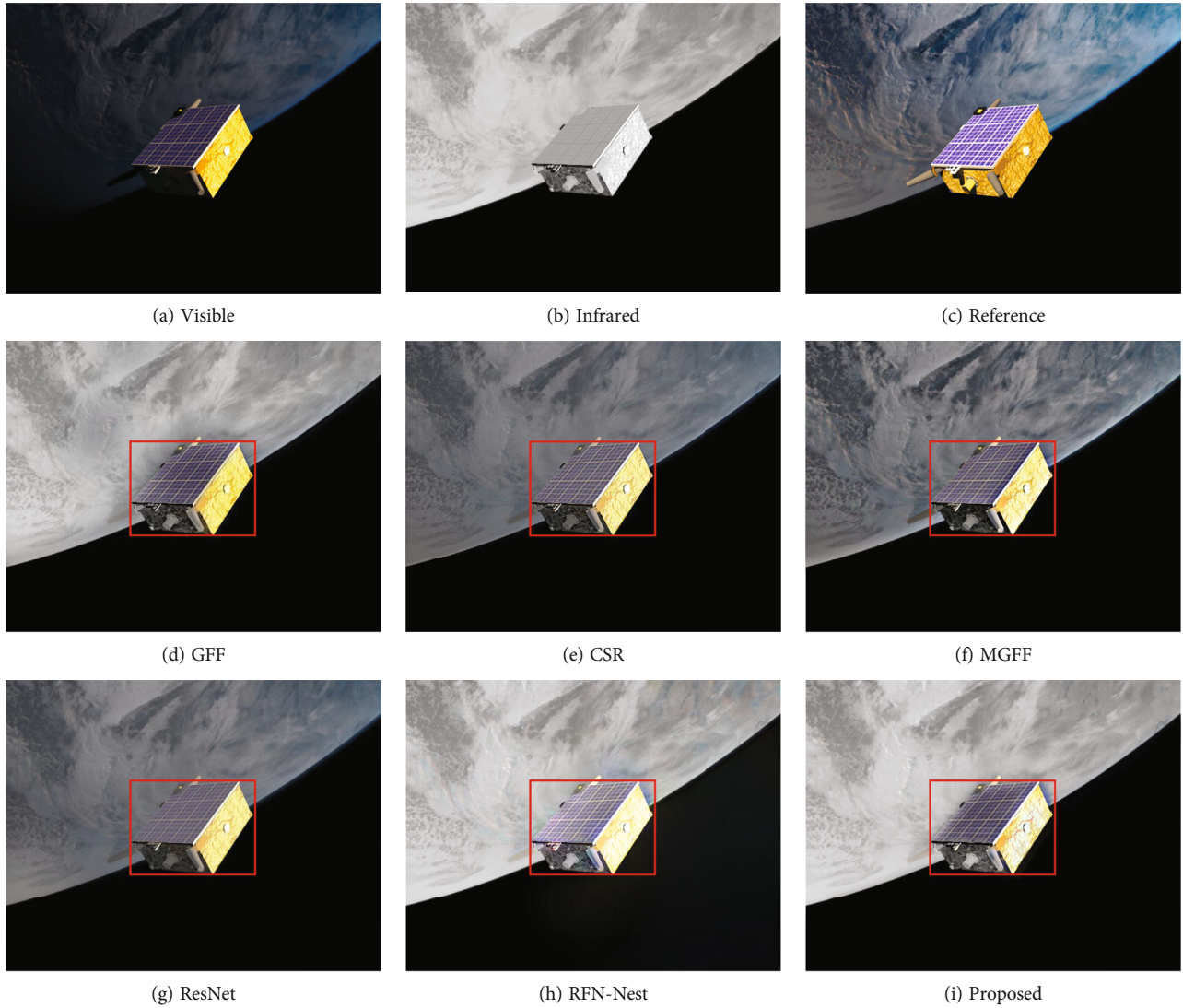


FIGURE 7: Fusion results of different methods in the weak illumination scene.

TABLE 3: Average evaluation metric values of all methods for local illumination scene of the SDD dataset. The best three values of each metric are marked as bold, italic, and underlined.

Method	AG	CE	EI	EN	MI	PSNR	$Q^{AB/F}$	$Q^{CB}$	RMSE	SD	VSI	SFF	GS
Proposed	1.7466	<u>0.4988</u>	18.0099	2.8169	<i>1.5896</i>	61.4582	<i>0.5001</i>	0.8434	0.0466	<b>74.7263</b>	<b>0.9372</b>	<b>0.8947</b>	<b>0.99</b>
MGFF	<b>2.0491</b>	<b>0.3948</b>	<b>20.948</b>	2.892	1.4816	<u>61.5836</u>	0.4747	0.833	<u>0.0453</u>	60.3331	0.8858	<u>0.7799</u>	0.9815
CSR	1.5921	0.566	16.2231	2.8255	<u>1.5848</u>	<i>61.7869</i>	0.4823	<u>0.8297</u>	<i>0.0433</i>	56.4709	0.8913	<i>0.7846</i>	<i>0.9844</i>
ResNet	1.1787	0.6066	12.44	2.8287	<b>1.6248</b>	<b>61.81</b>	0.3788	0.8162	<b>0.043</b>	56.2595	<i>0.9359</i>	0.768	<u>0.9828</u>
GFF	<u>1.812</u>	<i>0.4756</i>	<u>18.1636</u>	<i>2.9647</i>	1.5476	61.5493	<u>0.4991</u>	0.8229	0.0457	<u>69.5289</u>	<u>0.8951</u>	0.7384	0.9805
RFN-Nest	<i>1.8898</i>	1.0541	<i>19.1197</i>	<b>4.5598</b>	1.5663	61.3688	<b>0.5057</b>	0.4841	0.0479	<i>72.2098</i>	0.8937	0.6916	0.9794

TABLE 4: Average running time of all methods on a pair of images of size  $52 \times 582$  pixels.

Method	GFF	CSR	MGFF	ResNet	RFN-Nest	Proposed
Time (s)	0.67	5.46	1.77	9.61	4.65	2.62

overall performance by giving the 3 best values, 1 second best value, and 3 third best values. The ResNet method ranks the third place by giving the 3 best values, 1 second best value, and 1 third best value. Table 3 reveals that the proposed method obtains the largest value for VSI, SFF, and GS, which all are FR-IQA methods. It means that the fusion results of the proposed method are more closer to the

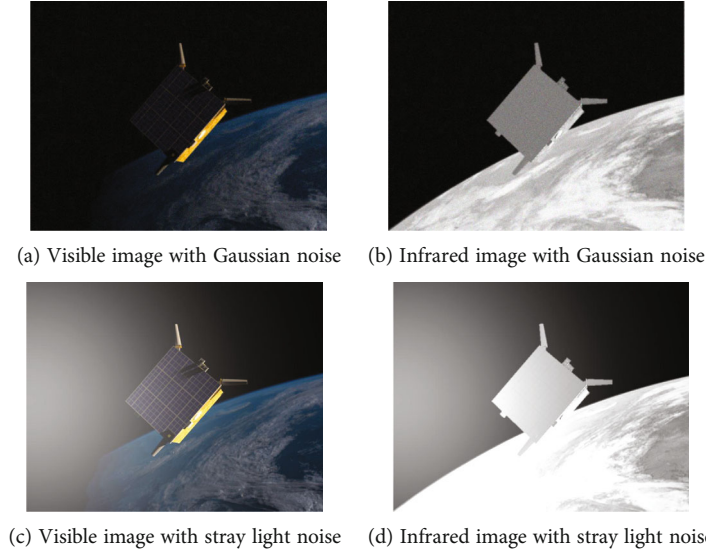


FIGURE 8: Example source images with Gaussian white noise and stray light noise.

TABLE 5: Average evaluation metric values of the MGFF method and the proposed method on Gaussian white noise of the SDD dataset. The best values of each metric are marked as bold.

Method	AG	CE	EI	EN	MI	PSNR	$Q^{AB/F}$	$Q_{CB}$	RMSE	SD	VSI	SFF	GS
Proposed	3.6143	<b>0.8873</b>	33.0336	<b>6.1675</b>	<b>2.0178</b>	59.1441	<b>0.4269</b>	0.4353	0.0793	<b>90.5489</b>	<b>0.9319</b>	<b>0.8425</b>	<b>0.9782</b>
MGFF	<b>4.6667</b>	1.2223	<b>42.6365</b>	6.1277	1.8367	<b>59.3035</b>	<b>0.3709</b>	<b>0.4949</b>	<b>0.0764</b>	59.8589	0.9059	0.7881	0.9714

TABLE 6: Average evaluation metric values of the MGFF method and the proposed method on stray light noise of the SDD dataset. The best values of each metric are marked as bold.

Method	AG	CE	EI	EN	MI	PSNR	$Q^{AB/F}$	$Q_{CB}$	RMSE	SD	VSI	SFF	GS
Proposed	1.9729	<b>0.5474</b>	19.969	<b>7.7421</b>	<b>6.0232</b>	59.5247	<b>0.6264</b>	0.626	0.0727	<b>78.043</b>	<b>0.9211</b>	<b>0.6845</b>	<b>0.9668</b>
MGFF	<b>2.4464</b>	0.9434	<b>24.7903</b>	7.5938	5.2711	<b>59.7492</b>	<b>0.6345</b>	<b>0.6425</b>	<b>0.069</b>	59.861	0.9113	0.3591	0.9655

reference image in spatial details. The MGFF method shows the largest value in AG and EI, which are feature-based metrics. This indicates that the MGFF method can well preserve the texture and edge information from the source images. The ResNet method gives the best performance in MI and PSNR, which denotes that it can well preserve the original information from source images. But it does not prove that the ResNet method can well combine the complementary information from the source images. Moreover, it is noteworthy that the deep learning-based method is not better than the traditional-based method.

**5.4. Computational Efficiency Comparison.** The computational efficiency of different image fusion methods over the image size of  $752 \times 582$  pixels is listed in Table 4. The proposed method and the CSR method are implemented in MATLAB integrated with Python. The Python part mainly computes the sparse coefficient maps for the convolutional basis pursuit denoising problem. This can be realized by the GPU-accelerated SPORCO package [53]. The dictionary filters adopted in the CSR method and the proposed method have the same settings. Both of them are learned from 50 natural image patches. The number of dictionary filter is

set as 32 and the size is set as  $8 \times 8$ . All the other methods are implemented in MATLAB. For the deep learning-based methods, the GPU is also utilized to generate the fused image. In Table 4, the conventional image fusion methods are faster than the deep-learning based methods. Although the proposed method is not as efficient as the GFF method and MGFF method, it is still feasible for space debris identification during space proximity operation missions. For example, it will take five to tens of minutes for the chaser to track the docking port during the station-keeping phase and final approach phase [54].

### 5.5. Robustness Analysis

**5.5.1. Noise.** The noise existing in the space surveillance platform mainly includes thermal noise, shot noise, dark current noise, and stray light noise. The former three noises are caused by the statistic nature of photodetection or photodiode in sensor systems [55], which can be represented as the Gaussian white noise model generally. In this paper, the noise with Gaussian blurring ( $\sigma = 1$ ) and zero-mean, Gaussian white noise ( $\text{var} = 0.0022$ ) are added to the source images referred to the SPEED datasets [56]. The stray light

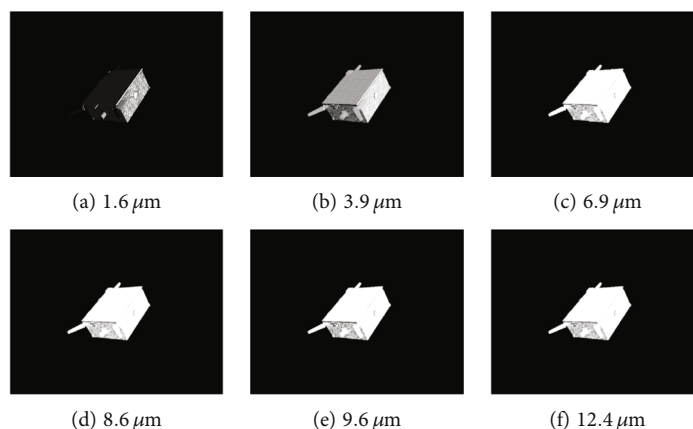


FIGURE 9: Example infrared images of different spectra.

TABLE 7: Average evaluation metric values of the proposed method of different spectra on the SDD dataset. The best three values of each metric are marked as bold, italic, and underlined.

Band	AG	CE	EI	EN	MI	PSNR	$Q^{AB/F}$	$Q_{CB}$	RMSE	SD	VSI	SFF	GS
1.6 $\mu\text{m}$	<b>1.2217</b>	7.0108	<b>12.394</b>	<b>2.508</b>	<b>0.589</b>	<b>55.2469</b>	<b>0.7137</b>	<b>0.0637</b>	<b>0.1943</b>	<b>49.4808</b>	<i>0.8542</i>	<i>0.8699</i>	<i>0.8955</i>
3.9 $\mu\text{m}$	<i>1.0079</i>	6.1274	<i>10.2168</i>	<i>2.433</i>	<i>0.5266</i>	<i>55.2148</i>	<i>0.672</i>	<i>0.0552</i>	<i>0.1957</i>	<i>32.8769</i>	<b>0.8717</b>	<b>0.8745</b>	<b>0.8968</b>
6.9 $\mu\text{m}$	<u>0.7747</u>	0.173	7.9118	<u>2.0605</u>	<u>0.2379</u>	55.1923	0.6392	0.0451	<u>0.1967</u>	<u>16.2434</u>	<u>0.8255</u>	<u>0.8651</u>	<u>0.8954</u>
8.6 $\mu\text{m}$	0.7413	<u>0.1678</u>	7.5526	2.0542	0.2277	55.1908	0.6338	0.0459	0.1968	15.5699	0.8236	0.8658	0.8952
9.6 $\mu\text{m}$	0.7354	<i>0.1464</i>	7.495	2.055	0.2273	55.1903	0.6321	0.0462	0.1968	15.4693	0.8236	0.8667	0.8953
12.4 $\mu\text{m}$	0.7201	<b>0.1397</b>	7.3307	2.0498	0.2205	55.1891	0.6297	0.0472	0.1969	15.0549	0.8222	0.8668	0.8952

noise results from the accidental perturbation of diffuse reflection from the Earth, moon, and other nebulas [57]. In particular, it will be even worse during Earth watching in LEO. It also can be modeled using two-dimension Gaussian function. The maximum gray value is set as 300, and the standard deviation is set as 300. The example visible images of space debris with Gaussian white noise and stray light noise are shown in Figure 8. The fusion results of the MGFF method and the proposed method are shown in Tables 5 and 6. As can be seen in Tables 5 and 6, the proposed method still outperforms the MGFF method within the same Gaussian white noise and stray light noise.

**5.5.2. Spectrum.** Six spectra of infrared images are synthesised, and these spectra distribute from near-infrared bands to far-infrared bands. The materials of the space debris are set by the material database in Vega software to reach the realistic infrared effects. Figure 9 shows the infrared images of different spectra. In Figure 9, the details drop drastically in higher bands and the second spectrum reserves the richest detail information. For excluding the influence of background, the source images of different spectra with a pure black background are analysed and the fusion results are listed in Table 7. We can see that the relationship between the spectrum and the fusion performance is not linear. Although the first spectrum reveals the best performance in almost all the fusion metrics, the second spectrum is thought to be the best fusion performance for it achieves

the best values in all the FR-IQA metrics. In conclusion, the second spectrum is suggested to be fused for space debris surveillance.

**5.5.3. Pose.** Owing to the characteristic of space debris tumbling, the pose of which varies greatly. Therefore, an investigation of the influence to the robust of different methods in different poses of space debris is meaningful in real space applications. Five pairs of source images with different pose are analysed, and the visible images are shown in Figure 10. The FR-IQA metrics of the proposed method and the MGFF method are shown in Figure 11. It can be seen that these two methods have the similar stability.

**5.5.4. Scale.** The scale of space debris in the image varies from the homing phase to the final approaching phase during typical scenes of space debris surveillance. Figure 12 shows five visible images in a different imaging range. Figure 13 shows the FR-IQA values of the proposed method and the MGFF method with the increase of the imaging range. Figure 13 still indicates similar stability of the two methods except for the VSI metrics.

**5.5.5. Feature extraction.** As described in Section 2.1, feature extraction is the key step of space debris identification techniques. In this section, we tested the robustness of feature extraction on the fused images in different illumination conditions. The KAZE [58] method is selected as the feature detector and feature descriptor for its superior performance. The different illumination conditions are considered as the

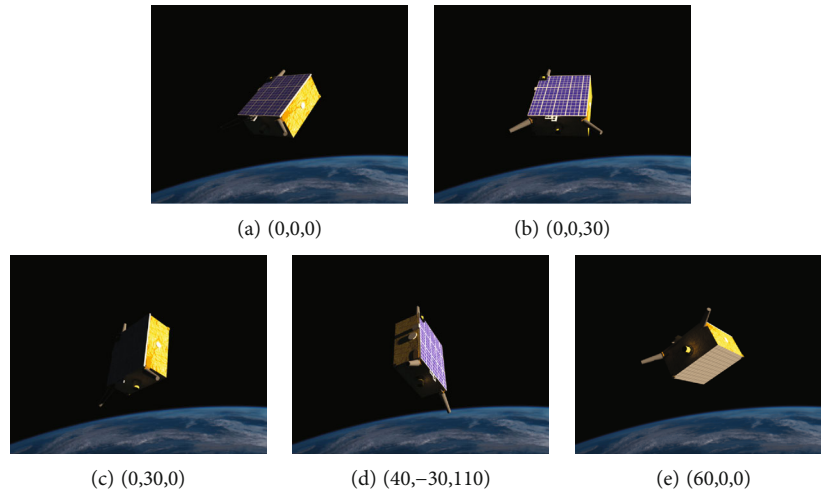


FIGURE 10: Visible images of different pose—unit (°).

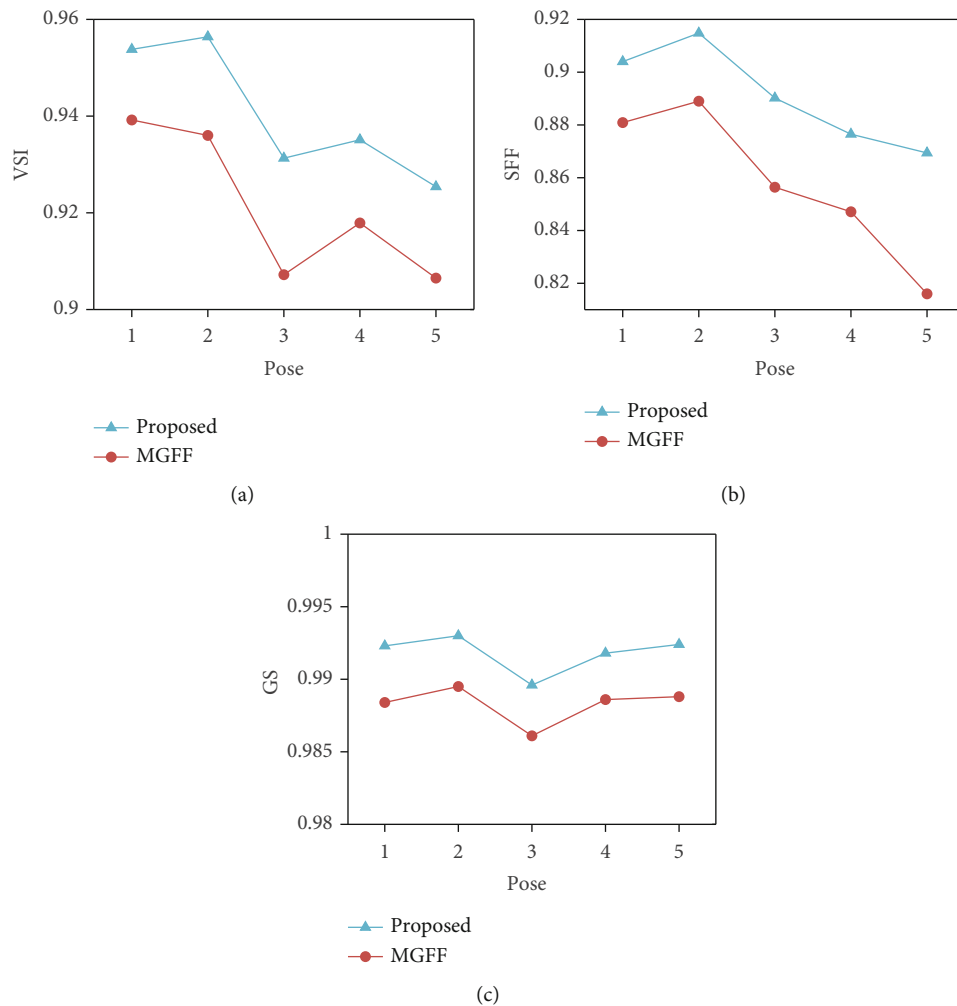


FIGURE 11: Fusion performance in different pose of space debris.

test scenarios. The illumination parameters are set based on the combination of linear and gamma transformation referred to [59]. The Y channel in the YCbCr color space is used as the metric of illumination intensity. The *matching*

*score* and *recall* versus *1-precision* curves are adopted as the robust evaluation of feature extraction. The *matching score* is defined as the ration between the number of correct matches and the smaller number of detected regions in the



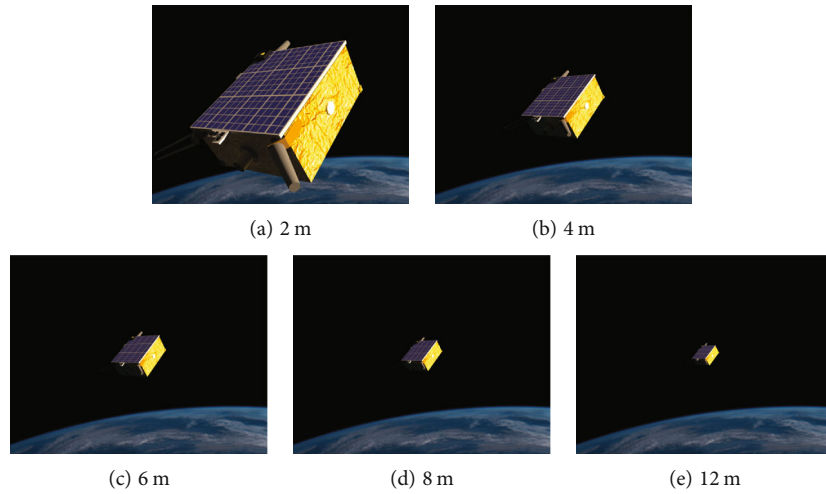


FIGURE 12: Visible images of different imaging range.

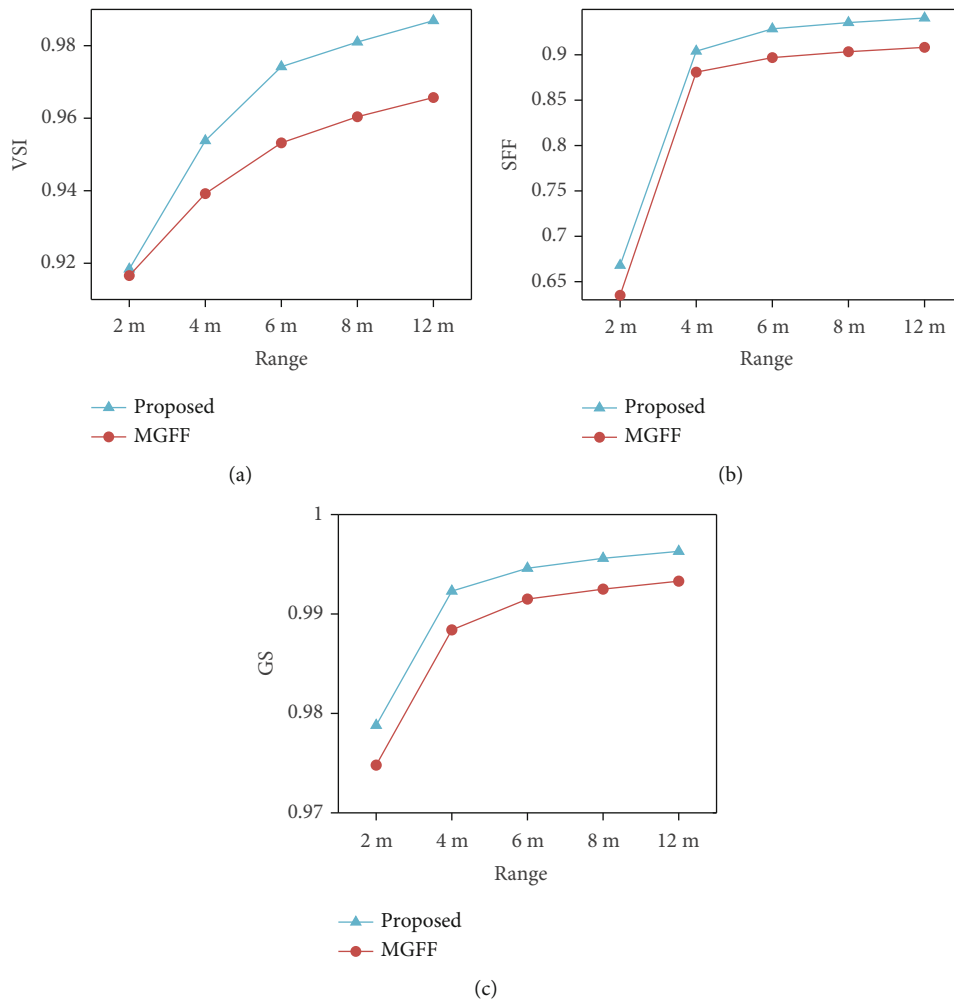


FIGURE 13: Fusion performance in different imaging range.

pair of images. The explicit definition of *recall* versus *1-precision* curves can be seen in [60]. The efficient approximate nearest neighbor search [61] is chosen as the matching method. The tested visible images with different illumination

parameters are shown in Figure 14, and the *matching score* and *recall* versus *1-precision* curves of the corresponding fused images are shown in Figures 15(a) and 15(b), respectively. As can be seen, the *matching score* and *recall* versus

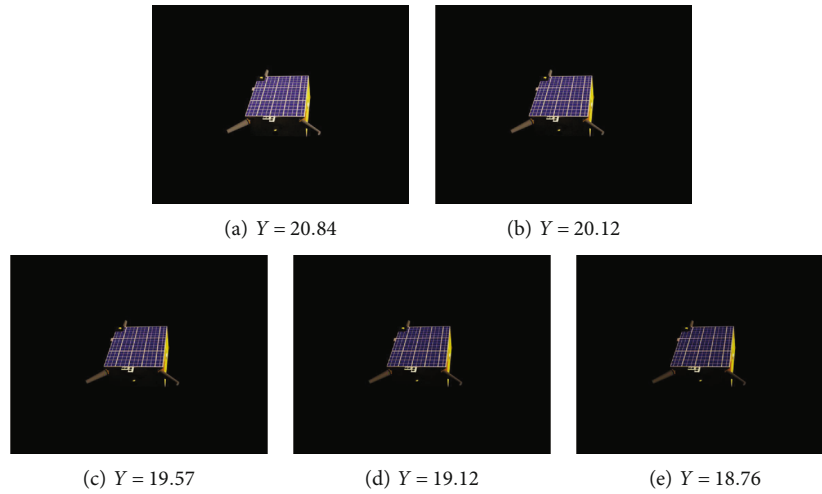


FIGURE 14: Fused images with different illumination conditions.

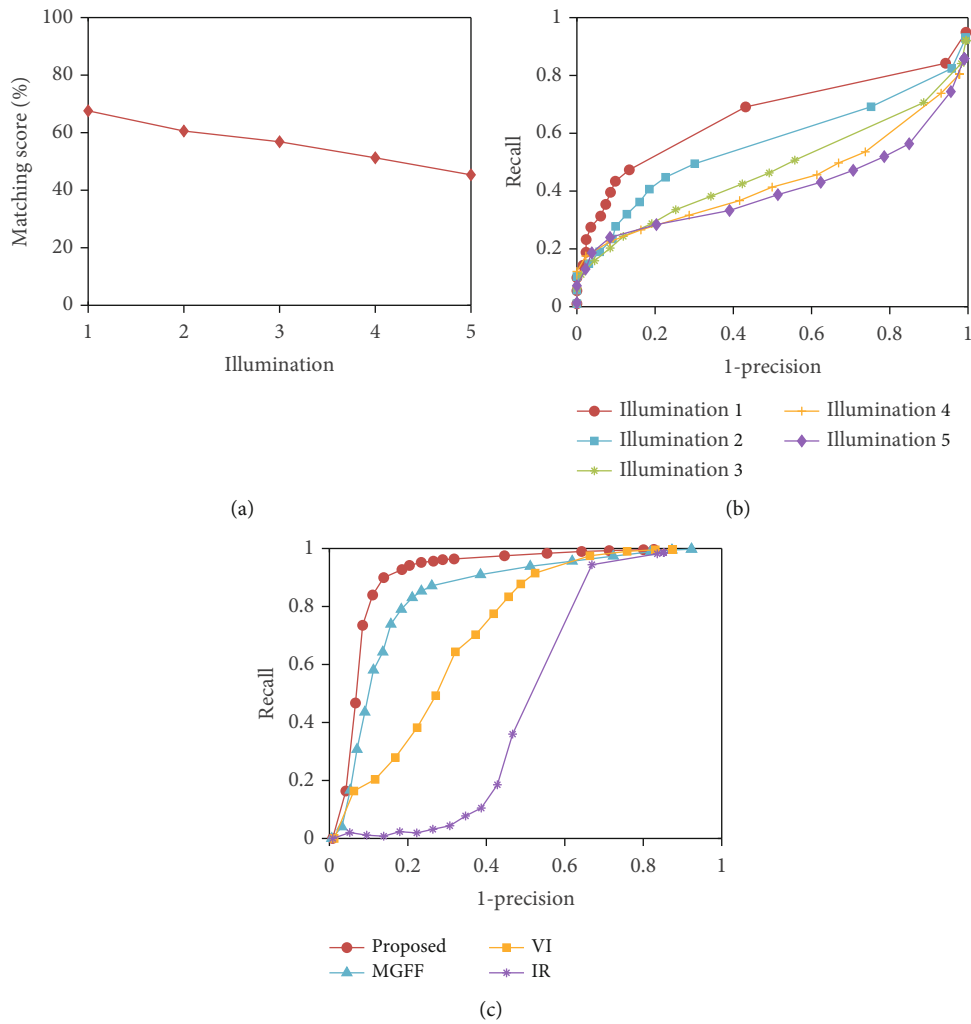


FIGURE 15: Robustness in different imaging range.

*1-precision* will decrease with lower illumination intensity. However, it still preserves a sufficient number of matches for the following object identification. Figure 15(c) depicted the feature extraction performance of fused images and

source images. It can be seen that the proposed fusion method possesses better robustness than the MGFF method. Both the fused images reveal higher feature extraction performance than the source images.

## 6. Conclusion

In this paper, we present an image quality enhancement framework for improving the on-orbit perception capability of the space debris surveillance platform, which is applicable for the weak illumination scene.

First, an improved guided filter is proposed to further enhance the space object and suppress the space noise. The improved guided filter utilizes local contrast measure based on the fact that the space object has a signature of discontinuity with its neighboring regions and can be considered as a homogeneous region.

Second, we propose a convolutional sparse representation-based image fusion method. The proposed method fuses the base layer by the guided filtering-based weighted strategy which makes full use of the strong correlations between neighborhood pixels, besides the detail layers fused by convolutional sparse representation-based method which can well preserve the details of the object.

More importantly, a public image fusion dataset for space debris surveillance is presented. To the best of our knowledge, this is the first public visible and infrared image fusion dataset in the field of space debris surveillance. The dataset contains six kinds of typical low Earth orbit scenes and four types of space debris.

We also test the robustness of the proposed method in different space noise, spectrum, pose, and imaging range. The experimental results demonstrate the advantages of the proposed method over other methods. The most appropriate spectrum of the infrared band for image quality enhancement is recommended. In the end, how to improve the fusion efficiency of the proposed method by accelerating the calculation of the CBPDN problem could be further investigated.

## Data Availability

The images data used to support the findings of this study have been deposited in the GitHub repository (<https://github.com/taojianggit/SDD>).

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the open project funds for the key laboratory of space photoelectric detection and perception (Nanjing University of Aeronautics and Astronautics), Ministry of Industry and Information Technology (no. NJ2020021-01), Fundamental Research Funds for the Central Universities (no. NJ2020021), and Funding for Outstanding Doctoral Dissertation in NUAA (no. BCXJ19-11).

## References

- [1] S. L. May, S. Gehly, B. Carter, and S. Flegel, "Space debris collision probability analysis for proposed global broadband constellations," *Acta Astronautica*, vol. 151, pp. 445–455, 2018.
- [2] Y. Fang, J. Pan, Y. Luo, and C. Li, "Effects of deorbit evolution on space-based pulse laser irradiating centimeter-scale space debris in LEO," *Acta Astronautica*, vol. 165, pp. 184–190, 2019.
- [3] O. Yilmaz, N. Aouf, L. Majewski, M. Sanchez-Gestido, and G. Ortega, "Using infrared based relative navigation for active debris removal," in *In: 10th International ESA Conference on Guidance, Navigation and Control Systems*, pp. 1–18, Salzburg, Austria, 2017.
- [4] J. Deloo and E. Mooij, "Active debris removal: aspects of trajectories, communication and illumination during final approach," *Acta Astronautica*, vol. 117, pp. 277–295, 2015.
- [5] G. B. Palmerini, "Combining thermal and visual imaging in spacecraft proximity operations," in *In: 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), IEEE*, pp. 383–388, Singapore, 2014.
- [6] S. Frank, S. Arne, J. Klaus, G. Manuel, and M. S. Gestido, "Lessons-learned from on-ground testing of image-based non cooperative rendezvous navigation with visible-spectrum and thermal infrared cameras," in *In: 10th International ESA Conference on Guidance, Navigation and Control Systems*, pp. 1–2, Salzburg, Austria, 2017.
- [7] S. Ivanov, B. Konstantinov, S. Tzokov et al., *Space Debris Identification, Classification and Aggregation with Optimized Satellite Swarms, Innovative Ideas for Micro/Nano-Satellite Mission*, Book Series 4, R. Sandau, R. Kawashima, S. Nakasuka, and J. J. Sellers, Eds., IAA Publ., International Academy of Astronautics, 2017.
- [8] M. Gang, J. Zhiguo, L. Zhengyi, Z. Haopeng, and Z. Danpei, "Full-viewpoint 3d space object recognition based on kernel locality preserving projections," *Chinese Journal of Aeronautics*, vol. 23, no. 5, pp. 563–572, 2010.
- [9] H. Pan, G. Xiao, and Z. Jing, "Feature-based image fusion scheme for satellite recognition," in *In: 2010 13th International Conference on Information Fusion*, pp. 1–6, Edinburgh, UK, 2010.
- [10] H. Ding, X. Li, and H. Zhao, "An approach for autonomous space object identification based on normalized ami and illumination invariant msa," *Acta Astronautica*, vol. 84, pp. 173–181, 2013.
- [11] J. Shi, J. Zhiguo, Z. Haopeng, and M. Gang, "Elastic net sparse coding-based space object recognition," *Acta Aeronautica et Astronautica Sinica*, vol. 34, no. 5, pp. 1129–1139, 2013.
- [12] H. Zhang and Z. Jiang, "Multi-view space object recognition and pose estimation based on kernel regression," *Chinese Journal of Aeronautics*, vol. 27, no. 5, pp. 1233–1241, 2014.
- [13] H. Zhang, Z. Jiang, and A. Elgammal, "Satellite recognition and pose estimation using homeomorphic manifold analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1, pp. 785–792, 2015.
- [14] H. Zhang, C. Zhang, Z. Jiang, Y. Yao, and G. Meng, "Vision-based satellite recognition and pose estimation using gaussian process regression," *International Journal of Aerospace Engineering*, vol. 2019, Article ID 5921246, 20 pages, 2019.
- [15] D. Zhao, M. Lu, X. Zhang, J. Shi, and Z. Jiang, "Satellite recognition via sparse coding based probabilistic latent semantic analysis," *International Journal of Humanoid Robotics*, vol. 11, no. 2, p. 1450008, 2014.

- [16] X. Yong, Z. Haoyue, and Z. Yanning, "Space Target Recognition Method Based on Deep Learning," in *In 2017 20th International Conference on Information Fusion (Fusion)*, pp. 1–5, Xi'an, China, 2017.
- [17] J. Tao, Y. Cao, M. Ding, and Z. Zhang, "Visible and infrared image fusion for space debris recognition with convolutional sparse representation," in *In: 2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC)*, pp. 1–5, Xiamen, China, 2018.
- [18] N. Dhamani, G. Martin, C. Schubert, P. Singh, N. Hatten, and M. R. Akella, "Applications of machine learning and monocular vision for autonomous on-orbit proximity operations," in *AIAA Scitech 2020 Forum*, p. 1376, Orlando, FL, 2020.
- [19] A. A. Grompone, *Vision-Based 3d Motion Estimation for On-Orbit Proximity Satellite Tracking and Navigation*, Naval Postgraduate School Monterey Ca, 2015.
- [20] S. Sharma, C. Beierle, and S. D'Amico, "Pose estimation for non-cooperative spacecraft rendezvous using convolutional neural networks," in *2018 IEEE Aerospace Conference*, pp. 1–12, Big Sky, MT, USA, 2018.
- [21] B. Chen, J. Cao, A. Parra, and T.-J. Chin, "Satellite pose estimation with deep landmark regression and nonlinear pose refinement," in *In: Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 2816–2824, Seoul, Korea, 2019.
- [22] W. D. Jongh, H. Jordaan, and C. V. Daalen, "Experiment for pose estimation of uncooperative space debris using stereo vision," *Acta Astronautica*, vol. 168, pp. 164–173, 2020.
- [23] S. Sonawani, R. Alimo, R. Detry, D. Jeong, A. Hess, and H. B. Amor, "Assistive relative pose estimation for on-orbit assembly using convolutional neural networks," *AIAA Scitech 2020 Forum*, vol. 1 PartF, 2020, pp. 1–11, AIAA, 2020.
- [24] R. Volpe, M. Sabatini, and G. Palmerini, "Shape reconstruction of a tumbling unknown orbital target by passive imaging," *Advances in the Astronautical Sciences*, vol. 170, pp. 15–29, 2020.
- [25] K. Kanani, A. Petit, E. Marchand, T. Chabot, and B. Gerber, "Vision based navigation for debris removal missions," in *63rd International Astronautical Congress*, Naples, Italy, 2012.
- [26] A. Petit, E. Marchand, and K. Kanani, "Vision-based detection and tracking for space navigation in a rendezvous context," in *In: Int. Symp. on Artificial Intelligence, Robotics and Automation in Space, i-SAIRAS*, Turin, Italy, 2012.
- [27] Q. Wei, Z. Jiang, and H. Zhang, "Robust spacecraft component detection in point clouds," *Sensors*, vol. 18, no. 4, p. 933, 2018.
- [28] Z. Haopeng, W. Quanmao, Z. Wei, W. Junfeng, and J. Zhiguo, "Sequential-image-based space object 3d reconstruction," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 42, no. 2, p. 273, 2016.
- [29] H. Zhang, Q. Wei, and Z. Jiang, "3d reconstruction of space objects from multi-views by a visible sensor," *Sensors*, vol. 17, no. 7, p. 1689, 2017.
- [30] M. Benn, *Vision Based Navigation Sensors for Spacecraft Rendezvous and Docking*, Technical University of Denmark, Kgs. Lyngby, Denmark, 2011.
- [31] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: recent advances and future prospects," *Information Fusion*, vol. 42, pp. 158–173, 2018.
- [32] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1823–1831, Santiago, Chile, 2015.
- [33] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1882–1886, 2016.
- [34] H. Li, X.-J. Wu, and J. Kittler, "Rfn-nest: an end-to-end residual fusion network for infrared and visible images," *Information Fusion*, vol. 73, pp. 72–86, 2021.
- [35] S. Singh and R. Anand, "Multimodal medical image fusion using hybrid layer decomposition with cnn-based feature mapping and structural clustering," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 3855–3865, 2020.
- [36] X. Jia, C. Zhu, M. Li, W. Tang, and W. Zhou, "Llvp: a visible-infrared paired dataset for low-light vision," in *In: Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3496–3504, 2021.
- [37] Y. Fu and X. Wu, "A dual-branch network for infrared and visible image fusion," in *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 10675–10680, Milan, Italy, 2021.
- [38] L. Tang, J. Yuan, and J. Ma, "Image fusion in the loop of high-level vision tasks: a semantic-aware real-time infrared and visible image fusion network," *Information Fusion*, vol. 82, pp. 28–42, 2022.
- [39] H. Zhang, H. Xu, X. Tian, J. Jiang, and J. Ma, "Image fusion meets deep learning: a survey and perspective," *Information Fusion*, vol. 76, pp. 323–336, 2021.
- [40] S. Sharma, S. D'Amico, S. Rock, and M. Schwager, *S. U. D. of Aeronautics & Astronautics, Pose Estimation of Uncooperative Spacecraft Using Monocular Vision and Deep Learning*, Stanford University, 2019.
- [41] P. Bodin, R. Noteborn, R. Larsson et al., "Prisma formation flying demonstrator: overview and conclusions from the nominal mission," *Advances in the Astronautical Sciences*, vol. 144, pp. 441–460, 2012.
- [42] M. Aviles, D. Mora, M. Canetri, and P. Colmenarejo, "A complete ip-based navigation solution for the approach and capture of active debris," in *In: 67th International Astronautical Congress*, Guadalajara, Mexico, 2016.
- [43] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2864–2875, 2013.
- [44] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [45] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 301–315, 2016.
- [46] T. V. Peters, D. Escorial, A. Pellacani, M. Lavagna, and M. A. Rodrigalvarez, "The cobra irides experiment," in *International Astronautical Congress*, pp. 1601–1611, Toronto, Canada, 2014.
- [47] L. P. Cassinis, R. Fonod, and E. Gill, "Review of the robustness and applicability of monocular pose estimation systems for relative navigation with an uncooperative spacecraft," *Progress in Aerospace Sciences*, vol. 110, article 100548, 2019.
- [48] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganieri, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative

- study,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 94–109, 2012.
- [49] G. Zhai and X. Min, “Perceptual image quality assessment: a survey, science in China series F,” *Information Sciences*, vol. 63, no. 11, article 211301, 2020.
- [50] D. P. Bavirisetti, G. Xiao, J. Zhao, R. Dhuli, and G. Liu, “Multi-scale guided image and video fusion: a fast and efficient approach,” *Circuits Systems and Signal Processing*, vol. 38, no. 12, pp. 5576–5605, 2019.
- [51] H. Li, X. J. Wu, and T. S. Durrani, “Infrared and visible image fusion with resnet and zero-phase component analysis,” *Infrared Physics & Technology*, vol. 102, article 103039, 2019.
- [52] X. Zhang, P. Ye, and G. Xiao, “Vifb: a visible and infrared image fusion benchmark,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 468–478, 2020.
- [53] B. Wohlberg, “Sporco: a python package for standard and convolutional sparse representations,” in *Proceedings of the 16th Python in Science Conference*, pp. 1–8, TX, USA, 2017.
- [54] D.-R. Lee and H. Pernicka, “Optimal control for proximity operations and docking,” *International Journal of Aeronautical and Space Sciences*, vol. 11, no. 3, pp. 206–220, 2010.
- [55] R. Hui, *Introduction to Fiber-Optic Communications*, Academic Press, 2020.
- [56] M. Kisantal, S. Sharma, T. H. Park, D. Izzo, M. Martens, and S. D’Amico, “Satellite pose estimation challenge: dataset, competition design and results,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 5, pp. 1–1, 2020.
- [57] J.-O. Park, W.-K. Jang, S.-H. Kim, H.-S. Jang, and S.-H. Lee, “Stray light analysis of high resolution camera for a low-earth-orbit satellite,” *Journal of the Optical Society of Korea*, vol. 15, no. 1, pp. 52–55, 2011.
- [58] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, “KAZE features,” in *Computer Vision – ECCV 2012. ECCV 2012. Lecture Notes in Computer Science, vol 7577*, Springer, Berlin, Heidelberg, 2012.
- [59] F. Lv, Y. Li, and F. Lu, “Attention guided low-light image enhancement with a large scale low-light simulation dataset,” *International Journal of Computer Vision*, vol. 129, no. 7, pp. 2175–2193, 2021.
- [60] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [61] M. Muja and D. G. Lowe, “Fast approximate nearest neighbors with automatic algorithm configuration,” *VISAPP (1)*, vol. 2, no. 331–340, p. 2, 2009.