



Article SFSNet: An Inherent Feature Segmentation Method for Ground Testing of Spacecraft

Yingxiao Li , Ping Ma and Ju Huo *

Control and Simulation Center, Harbin Institute of Technology, Harbin 150080, China; hitliyingxiao@163.com (Y.L.); pingma@hit.edu.cn (P.M.)

* Correspondence: torchjuhuo@163.com

Abstract: In the final approach stage of rendezvous and docking of a spacecraft, the pose parameters of the target spacecraft need to meet docking or berthing capture conditions. Visible light visual measurement systems are increasingly employed in spacecraft ground tests to extract the geometric features of spacecraft to calculate and verify the accuracy of pose parameters. Most current featuresegmentation algorithms are unable to break through the scale transformation problem of spacecraft movement and the noise interference of multi-layer insulation materials in imaging. To overcome these challenges, we propose a novel feature segmentation algorithm based on the framework of deep convolutional neural networks. Firstly, a full convolution model of the encoding-decoding structure is constructed based on data for the ground test. A feature concatenation module is applied and combined with a network backbone to improve the segmentation performance. Then, a comprehensive loss function is presented and optimized by the pose characteristics of the spacecraft in the approach phase. Furthermore, a specific spacecraft simulation dataset to train and test our segmentation model is built through data augmentation. The experimental results verify that the proposed method achieves accurate segmentation of spacecraft of different scales, suppresses the interference caused by multilayer insulation material, and has strong robustness against motion ambiguity. The pixel accuracy of our proposed method reaches 96.5%, and the mean intersection over union is 93.0%.

check for updates

Citation: Li, Y.; Ma, P.; Huo, J. SFSNet: An Inherent Feature Segmentation Method for Ground Testing of Spacecraft. *Aerospace* 2023, *10*, 877. https://doi.org/10.3390/ aerospace10100877

Academic Editor: Pierre Rochus

Received: 7 August 2023 Revised: 17 September 2023 Accepted: 27 September 2023 Published: 11 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). **Keywords:** ground test; feature segmentation; convolution network; data augmentation; small sample learning

1. Introduction

As a comprehensive simulation test for on-orbit services, spacecraft ground testing ensures stable operation in their designated orbits. In ground tests that simulate closerange rendezvous operations, the position and pose parameters of the target spacecraft are critical factors. The accuracy of these parameters aids in determining the feasibility of capturing the target spacecraft through docking or berthing mechanisms. Typically, the propulsion system's drive motor can transmit its motion parameters to the control terminal. However, due to inherent hardware system errors, auxiliary measurement methods have become increasingly necessary in recent years [1,2] to validate the reliability of these parameters. Moreover, the technique of multi-sensor data fusion measurement has emerged as a fundamental technology for autonomous rendezvous and docking in modern spacecraft. By integrating data from various sensors and employing advanced fusion algorithms, this technology enhances the accuracy and robustness of the measurements, ultimately ensuring the safe and successful completion of these intricate and high-stakes missions.

As a direct approach for auxiliary measurement, computer vision-based technology has been increasingly vital in spacecraft pose estimation. A stereo vision measurement system [3] is not only suitable for auxiliary measurement tasks, but also possesses the advantages of being non-contact, having low power consumption, and operating at low cost [4,5], making intelligent in-orbit services more feasible and efficient.

During ground testing, the measurement process of visual pose estimation is illustrated in Figure 1. Initially, the target spacecraft adheres to its on-orbit operational parameters while situated within the field of view (FOV) of several stationary stereo camera pairs, which are employed for image data acquisition [6]. Next, feature extraction algorithms [7,8] obtain the spacecraft's feature data from the collected images. Simultaneously, camera calibration [9] calculates the spatial geometric relationship between the pixel plane, image plane, and stereo cameras, determining their intrinsic and extrinsic parameters. Finally, the extracted feature information is combined with the calibration parameters to estimate the spacecraft's six degrees of freedom (6-DoF) pose relative to the ground, completing the pose estimation process.



Figure 1. Procession of spacecraft vision pose estimation.

As depicted in Figure 1, the precision of the extracted spacecraft features has a great impact on the accuracy of subsequent pose calculations. For cooperative spacecraft, their distinctive markers are easily distinguishable from other regions in the image. Commonly, these markers can be identified using basic algorithms, such as feature detection operators [10] and adaptive thresholding methods [11]. However, in most real spacecraft docking scenarios, there are no cooperative signals at the visual image level. To realistically simulate the approach process, features must be extracted from non-cooperative targets. In recent applications, an effective method for feature extraction is identifying inherent geometric shapes or feature clusters of the spacecraft, such as the regular contours of the spacecraft's main body or the solar panel brackets. Miao et al. [12] combined canny detection with the geometric constraints of the satellite and determined its outer rectangular contour. Peng et al. [13] applied a 4-adjacency pixel based sliding window method and extracted the circular features of the docking rings. Zhang et al. [14] merged similar line segments that were detected by Hough transform, and obtained the true triangle bracket edge using distance and area constraints. Liu et al. [15] obtained the multi-elliptical features of the nozzle of a spacecraft using an improved bee colony algorithm. Another feasible method utilizes the relatively invariant local feature clusters in an image as the target feature. Longge et al. [16] extracted the feature point clusters of a spacecraft by applying the Otsu threshold and using an improved FAST corner algorithm. Huang et al. compared SIFT and its improved algorithm to obtain the feature sets of the target vehicle. Wei et al. [17] segmented the basic contour of the target satellite by applying a weak gradient elimination method, but it also weakened the geometric information of the spacecraft and was not able to distinguish the texture and the edge of the spacecraft.

However, the image data used in the aforementioned algorithms are idealized simulation images generated by modeling software, and the texture information of these images is very different from real shot images from a ground test. Consequently, these methods struggle to be applied in practical engineering applications and to extract the target's features directly. Specifically, there are three challenges that need to be addressed:

- (1) In order to simulate the rendezvous stage, the spacecraft moves toward or away from the camera, which, accordingly, changes its scale within the FOV. This variation between consecutive frames undermines detection algorithms that rely on the target's full size and the detection distance condition, causing them to lose their prior constraints. The spacecraft's position in the camera's FOV and the size of the receptive field it occupies are constantly changing, which greatly impacts the constraints in traditional feature extraction methods;
- (2) With the advancement of camera manufacturing, industrial cameras have rapidly improved in performance, particularly in resolution. Higher pixel counts can significantly enhance image detection accuracy. However, as image resolution increases, traditional visual algorithms, such as the sliding window method, incur a significant increase in computational costs and processing time when traversing features;
- (3) In the field of contemporary space technology, spacecraft surfaces are frequently coated with polyimide (PI) multilayer insulation material (MLI), as depicted in Figure 2. This material enables the spacecraft to withstand extreme temperature variations in space and ensures the stable operation of its electronic components. The texture of MLI presents a prominent feature in gray-scale images, which significantly hinders the extraction of pixel-level feature edges and ultimately affects the detection performance of traditional feature extraction algorithms.



Figure 2. Polyimide (PI) multilayer insulation material and its applications in the space industry. Generally, MLI, as shown in (**a**), is composed of PI and aluminum foil. (**b**,**c**) depict examples of MLI film applications on the Apollo lunar module and the Mangalyaan Mars probe, respectively.

Fundamentally, the feature extraction task for a non-cooperative target involves classifying different pixel categories and clarifying their boundaries. Traditional vision algorithms typically associate local pixels and employ specific filtering or clustering operators to classify pixels and form identifiable features. However, a limitation of these methods is their reliance on low-level pixel semantic information, which may result in local optimal solutions when encountering challenges akin to those mentioned in (2).

With advancements in computational power, deep learning methods have demonstrated their superiority in the field of computer vision. To solve the issues above and extract the features of spacecraft accurately, a spacecraft feature segmentation network (SF-SNet) is proposed in this article, which is a deep learning algorithm based on convolutional neural networks (CNNs). As a hierarchical structure method, two-dimensional convolution can generate feature maps of different scales, capturing more details and utilizing greater image information compared to traditional characteristic operators.

The remainder of this paper is structured as follows: Section 2 describes the proposed spacecraft feature segmentation algorithm. Section 3 analyzes the experimental results and compares them under various conditions. Section 4 summarizes the main achievements of this study and presents the outlook for future spacecraft segmentation tasks.

2. Spacecraft Feature Segmentation Network Model

In this section, the spacecraft feature segmentation model is presented in the context of ground test applications. Firstly, based on the size of the test image, a fully convolutional network is employed to construct the encoder–decoder backbone structure, which classifies pixels belonging to different categories in the image. Moreover, to enhance the transmission and utilization of information under different receptive fields, a concentrating module is used to connect the encoding and decoding stages of the backbone. Lastly, considering the geometric characteristics of spacecraft, a pixel clustering loss function is established to ensure the classification accuracy and to reduce the training complexity of the model.

2.1. Basic Module and Encoder–Decoder Structure

In order to classify each pixel in the image as accurately as possible and to minimize the positioning loss of spacecraft features, an encoder–decoder segmentation structure [18,19] is applied to construct the primary architecture of SFSNet. This type of structure not only preserves the spatial information of the input data, but also implements pixel-level dense estimation tasks.

In the encoding stage, image data is transformed into high-dimensional space vectors via nonlinear mapping. In the decoding stage, the generated vector is utilized to restore the spatial dimensions and detailed information of the original image. Data dimension transformation is accomplished through convolution, up-sampling, and pooling operations.

Convolution processing in the image task involves discrete convolution computation. Spacecraft images are filtered using various convolution kernels to obtain diverse spacecraft features, thereby constructing a feature map rich in information. This advantage is unparalleled compared to the single-feature filter in divide and conquer algorithms, such as canny or Laplacian operators [20,21]. The formal convolution operation can be expressed as the Equation (1):

$$\boldsymbol{y}_{i^{c+1},j^{c+1},d} = \sum_{i=0}^{H} \sum_{j=0}^{W} \sum_{d^{c}=0}^{D^{c}} \boldsymbol{f}_{i,j,d^{c},d} \times \boldsymbol{x}_{i^{c+1}+i,j^{c+1}+j,d^{c}}^{c}$$
(1)

where $x^c \in \mathbb{R}^{H^c \times W^c \times D^c}$, $y^c \in \mathbb{R}^{H^c \times W^c \times D^c}$ are the input and output tensors of the convolution layer *c*, respectively. $f^c \in \mathbb{R}^{H \times W \times D^c}$ are the convolution kernels of *c*. It follows that convolution kernels of the same layer share their weights. *H*, *W*, *D* are the height, width and channel number of the corresponding tensor. *i*, *j*, *d* is the position of the convolution result, which is constrained by the Equation (2):

$$\begin{cases} 0 \le i^{c+1} < H^{c+1} \\ 0 \le j^{c+1} < W^{c+1} \end{cases}$$
(2)

Furthermore, to obtain accurate spacecraft features, data from higher receptive fields [22] are required. Pooling layers are employed in this paper to modify image dimensions and to acquire feature data under various receptive fields, enabling the acquisition of high-level semantic information. To achieve rapid and efficient feature extraction, the maximum pooling method is utilized to reduce the image dimensions in the encoding stage, as demonstrated in Equations (3) and (4).

$$\mathbf{y}_{i^{c+1},j^{c+1},d} = \max_{0 \le i \le H, 0 \le j \le W} x_{i^{c+1} \times H+i,j^{c+1} \times W+j,d^c}^{c}$$
(3)

$$\begin{cases}
0 \le i^{c+1} < H^{c+1} \\
0 \le j^{c+1} < W^{c+1} \\
0 \le d < D^{c+1} = D^c
\end{cases}$$
(4)

In the decoding stage, up-sampling layers are deployed as the inverse operation of pooling layers to reconstruct the dimensions of the segmented image. As demonstrated

in Equation (5), the nearest neighbor interpolation method is used in the up-sample layer, offering a fast and efficient approach for restoring the scale of spacecraft images.

$$\begin{cases} i^{c} = i^{c+1} \times (W^{c}/W^{c+1}) \\ j^{c} = j^{c+1} \times (H^{c}/H^{c+1}) \end{cases}$$
(5)

2.2. Feature Concatenation Module

Within the SFSNet backbone, the input image's dimensions are reduced during the encoding phase. While this assists the model in perceiving the image's semantic information, with deepening of the network, the height and width of the feature map gradually decrease; thus, the resolution of the spacecraft continuously decreases, resulting in the loss of a large quantity of detailed data of the original feature.

The up-sampling layer in the decoding phase cannot provide additional data for image restoration. Consequently, the segmentation error in the spacecraft's edge or corner areas is significant, or these features may not even be extracted. To address this issue, a feature concatenation module is implemented between the encoding and decoding stages, enhancing the ability of the underlying pixel information to propagate backwards, as illustrated in Equation (6) and Figure 3.

$$\boldsymbol{y}_{m}^{H \times W \times (D_{1} + D_{2})} = \operatorname{concat}\{\boldsymbol{y}_{1}^{H \times W \times D_{1}}, \boldsymbol{y}_{2}^{H \times W \times D_{2}}\}$$
(6)

First, the feature map for restoration, $y_2 \in \mathbb{R}^{H \times W \times D_2}$, is generated through the upsampling layer. Next, the feature map, $y_1 \in \mathbb{R}^{H \times W \times D_1}$, which has the same width and height as y_2 in the encoding stage, is merged with y_2 in the channel direction to form the merge layer. Finally, the new feature map is up-sampled through channel dimension reduction, and the subsequent merged layer is recursively constructed.



Figure 3. Structure schematic of the feature concatenation module. In the actual model construction, a convolution layer is required between the merged layer and the subsequent up-sampling layer to reduce the channel dimension of the former.

The advantage of constructing the merged layer using the concatenate module is that this retains more valuable features from the original spacecraft image. These features can participate in the model training through standalone convolution parameters. By incorporating the feature concatenation modules into multiple levels of the SFSNet's backbone, details of the spacecraft at various scales are added to the restoration stage, thereby improving the precision of the model's segmentation results.

2.3. Spacecraft Characteristic Objective Function

After constructing the feed-forward propagation of the SFSNet, an appropriate feature segmentation objective function is required to determine the spacecraft pixel positioning error, thereby establishing the complete error back-propagation route. As a multisegmentation task mentioned in Section 1, the exterior contour and the circular feature should be segmented separately from the background. Generally, the softmax function and the cross-entropy loss function are effective for classifying pixels in multi-classification tasks. As shown in Equations (7) and (8), the softmax function connects to the last layer of SFSNet, converting the output data into pixel category probabilities and transmitting them to the cross-entropy function.

$$p_{i,j,d} = S(y_{i,j,d}) = \exp((y_{i,j,d})) / \sum_{c=0}^{C} \exp((y_{i,j,d}^{c}))$$
(7)

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} b_i^c \log(p_i^c)$$
(8)

where *S* represents the softmax function, and *L* represents the cross-entropy loss function. $p_{i,j,d}$ represents the pixel category probability under the image coordinates (i, j, d), while *C* represents the preset number of categories of the image pixel clusters. For the gray-scale image and the multi-classification task in the ground test, d = 1, C = 3. *N* represent the total number of pixels in the output image. b_i^c is a Boolean function, which equals 1 if the pixel category of the ground truth of a pixel is consistent with the pixel category of the model output, and 0 otherwise.

During the final approach phase of close-range rendezvous operations, the side of the spacecraft equipped with a docking or berthing mechanism consistently faces the measurement vision system, with the spatial angle between the axes of their roll direction being extremely small. Given that the spacecraft's main body and the rendezvous structure are non-deformable mechanical structures, their relative positions remain constant. Moreover, the ring-type berthing mechanism is consistently situated within the exterior contour of the spacecraft. Therefore, the multi-class segmentation task can be conceptualized as a task of separating the enclosed area, which is constituted by the internal region of the spacecraft's exterior contour and the external region of the circular berthing ring, as shown in Figure 4.

$$p_{i,j,d} = S(y_{i,j,d}) = 1/(1 + \exp(y_{i,j,d}))$$
(9)

$$L = -\frac{1}{N} \sum_{i=1}^{N} b_i \cdot \log p_i + (1 - b_i) \cdot \log(1 - p_i)$$
(10)

Consequently, the spacecraft segmentation task presented in this article can be regarded as a binary segmentation problem. In this context, the loss function is represented by a sigmoid function and a binary cross-entropy function, as delineated in the Equations (9) and (10). It can be inferred that Equations (9) and (10) are simplifications derived from Equations (7) and (8), respectively. For binary classification problems, p_i and $1 - p_i$ are the only two probability distributions. This simplified objective function enhances the computational efficiency and expedites the convergence of SFSNet.



Figure 4. The multi-class segmentation task is reformulated into a binary segmentation problem, segregating pixels within the blue region from those outside. The magnified images on either side illustrate the stringent demand for pixel-level accuracy in the annotation mask.

2.4. Overall Network Model of SFSNet

In the overall model construction, the encoder-decoder structure serves as the backbone of the proposed model.

$$LeakyReLU(x) = \begin{cases} \alpha \cdot x & x < 0\\ x & x \ge 0 \end{cases}$$
(11)

Table 1 depicts the basic operation layers used in various modular constructions, where 'Conv' denotes a convolution layer. All of the Conv layers are activated by the Leaky-ReLU [23] function, as directed in Equation (11). The function of α is that, in the process of back-propagation, where the Leaky-ReLU activation function input is less than zero, the gradient can also be calculated to avoid the problem of neuron death. According to [23] and prior testing, the hyper-parameter α is set as 0.1 in this article. 'Maxpool' signifies a maximum pooling layer, and 'Upsample' refers to the nearest neighbor interpolation layer.

Module	Type/Stride	Filter Param	
CPM1	Conv/(1,1)	3×3	
	Conv/(1,1)	3×3	
	Maxpool/(2,2)	2×2	
CPM2	Conv/(1,1)	3×3	
	Conv/(1,1)	3×3	
	Dropout/-	0.5	
	Maxpool/(2,2)	2×2	
UMCM	UpSample	2×2	
	Conv/(1,1)	2×2	
	Merge/-	channel = -1	
	Conv/(1,1)	$3 \times 3 s$	
	Conv/(1,1)	$3 \times 3 s$	

In addition, to further increase the nonlinear mapping capability of SFSNet and to prevent overfitting [24], 'dropout' layers are deployed in the encoding stage. 'Merge' stands for the feature concatenation module. Table 2 delineates the entire model's architecture and the output parameters associated with each modular component. In total, SFSNet has 33 convolution layers, 3 dropout layers, 6 pooling layers, 6 up-sampling layers, and 6 feature concatenation modules. The total number of optimized parameters in SFSNet is 10.9×10^{6} .

Table 2. Structure of backbone.

Phase	Module or Layer	Module or Layer
	CPM1	(512,512,16)
	CPM1	(256,256,32)
Encoding	CPM1	(128,128,64)
Encoding	CPM2	(64,64,128)
	CPM2	(32,32,256)
	CPM2	(16,16,512)
	Conv	
Linkage	Conv	(16,16,1024)
0	Dropout	
Decoding	UMCM	(32,32,512)
	UMCM	(64,64,256)
	UMCM	(128,128,128)
	UMCM	(256,256,64)
	UMCM	(512,512,32)
	UMCM	(1024,1024,16)
	Conv	(1024,1024,1)

3. Experiment Preparation and Performance Analyses

This section begins with an explanation of the dataset construction, which includes the process of image data collection and the deployment of augmentation strategies. Subsequently, to better adapt the model to the dataset and corresponding application scenarios, the segmentation efficiency under different modular structures and training strategies is compared and analyzed.

The experiments presented in this paper were conducted on an Intel Core i7 9700 3.0 GHz processor and an NVIDIA 2070s graphics card. To enhance the model's performance, the system memory was upgraded to 48 GB, thereby increasing the efficiency of the image generator and the size of the mini-batch. The experiments utilized the Windows 10 operating system and were conducted in the PyCharm Integrated Development Environment (IDE) with Python 3.6. The supporting frameworks included CUDA 9.0, cuDNN 7.3.0, and Tensorflow 1.8.0.

3.1. Image Data Collection and Augmentation

3.1.1. Dataset and Mask Acquisition

As a data-driven classification model, the parameters and performance of the SFSNet are largely determined by the image data and corresponding masks. Distinct from image classification or detection tasks, semantic segmentation masks demand superior labeling precision for achieving pixel-level accurate segmentation. An accurately labeled image dataset is crucial for training a segmentation model. To be specific, the ground test issues referred to in the Introduction are elaborated in the following:

- (1) At present, public datasets are frequently employed in the field of computer vision. Nevertheless, the application scenarios and target objects of these datasets, such as COCO and KITTI, are quite disparate from the data requirements of spacecraft ground testing. Hence, it is inefficient to use data from these datasets for enhancing the segmentation performance of SFSNet through instance-based transfer learning [25]. Thus, considering both the training cost and the segmentation performance, spacecraft images were captured and used as the primary data for the dataset in this paper. The ground test scene and experimental process were simulated, with a model spacecraft used for image data capture. The simulated experimental process can be substituted with a practical ground test task.
- (2) The purpose of the vision-based spacecraft ground test is typically to validate its pose and motion parameters. When the spacecraft moves along the camera's depth of field, this motion manifests as a change in the spacecraft's scale within the image sequence. This issue is a significant impediment to the efficiency of conventional feature extraction algorithms. Therefore, it is necessary to include image data from various depth-of-field positions of the spacecraft in the dataset to equip SFSNet for this measurement requirement.
- (3) Due to the spacecraft's motion and focusing sensitivity constraints, the collected video data occasionally exhibited blurring during the experiment. Given the spacecraft's slow motion, the distortion commonly manifests as a defocus blur, as illustrated in Figure 5c. Defocus blur can impact the model's segmentation performance within this image sequence. Therefore, to enhance the model's robustness, it is necessary to incorporate data exhibiting different levels of defocus blur into the training set. This operation also serves to prevent model overfitting. Consequently, the mask of a clear image taken under the same conditions is used as the label for the distorted image.

3.1.2. Single Sample Data Augmentation

As mentioned above, the precision of spacecraft feature semantic segmentation should attain pixel-level accuracy, necessitating stringent requirements for label masks. As a supervised learning model based on a brand new dataset, the data mask of SFSNet requires manual annotation. It is evident from Figure 4 that the manual annotation method demands



significant effort to ensure pixel-level accuracy in the mask, and this workload sharply escalates as the image resolution increases.

Figure 5. Clear image sample (**b**) and corresponding distortion scenarios. During the target's movement, data distortion mainly comprises of motion blur (**a**) and defocus blur (**c**). Typically, in ground tests, the target spacecraft does not undergo abrupt tangential motion; hence, the primary source of noise is the defocus blur.

To reduce the time and effort invested in annotating, a data augmentation technique, which is tailored to the specificities of the spacecraft, is employed in this paper. Specifically, considering the spacecraft as a non-deformable rigid body, its motion does not alter the relative positional relationships within its structure. Similarly, operations like rigid transformations of image data do not disturb the correlation among the pixels of the spacecraft body within the field of view. However, these preprocessed images are brand new to the segmentation network, and they can exert varied influences on the network parameters compared to the original data [26]. This approach enhances the robustness of the SFSNet model, particularly in contexts where the original data is scarce.

Considering the unique conditions of the ground test environment and the specific characteristics of the captured photos, single sample data augmentation is adopted in this paper. This method augments data through operations such as flipping, translating, and defocus, etc. For example, when performing rotation augmentation, a rotation angle is randomly generated according to the rotation range in Table 3. Rotating the image to be trained and its ground truth mask at the same angle brings them into the network training as new training data. Single-sample augmentation not only effectively generates potential spacecraft imaging scenarios, but also boasts the advantages of low complexity and straightforward implementation.

Moreover, given the low brightness of background pixels in gray-scale images from the tests, which are easily identifiable, the augmented image's missing pixels are filled with constant-value ones in this paper. This ensures that the enhanced data retain a resemblance to the original image.

Additionally, semi-supervised or unsupervised learning augmentation methods, which can automatically generate label masks (such as generative adversarial networks [27]), offer greater strength than simple single-sample augmentation. However, considering these algorithms are typically designed for models with extensive original data volumes and require high-end computing hardware, they are currently not feasible for the SFSNet.

In conclusion, considering our GPU's computational capacity and the pixel count of recent satellite cameras [28,29], the original simulation spacecraft dataset comprised 75 images with a resolution of 1024×1024 . Among them, 50 images were regarded as the original image set for data augmentation, 15 were used as the verification set to observe the model trend, and 10 were applied as the test set to evaluate the final model. Initially, all images were annotated with ground truth masks. Subsequently, based on the described augmentation method, the parameters range for the image generator were set as illustrated in Table 3, and the training set images and their masks were fed into the generator for random transformation.

Augment Type	Param Range
Rotation	$0 \sim 180^\circ$
Horizontal Translation	$0{\sim}0.5^{-1}$
Vertical Translation	$0{\sim}0.5$
Zoom	0~0.3 ²
Horizontal Flip	random boolean
Vertical Flip	random boolean
Defocus Blur	0~10 ³

Table 3. Data augmentation parameters.

_

¹ The ratio of the translation pixels to the total width/height pixels of the image; ² Image zoom scale is [1 - param, 1 + param]; ³ The radius of the defocus blur filter, in pixels.

3.2. Experiment Condition and Training Strategy

Before using augmented data to train SFSNet, the weight matrix of the model needs to be initialized. The Kaiming normal distribution initializer [30] was deployed in the initialization procession of SFSNet, in order to prevent the data gradient from exploding or disappearing in the forward propagation process. In addition, at the start of training, L_2 -normalization was added in the convolution kernel regularization procession to avoid overfitting of the network, which is given by Equation (12):

$$L_2 = \frac{1}{2} \cdot \lambda \|\omega\|_2^2 = \lambda \sum_j \omega_j^2$$
(12)

where ω are the parameters of the weight matrix, λ is the L₂-norm parameter, which is a regularized hyper-parameter that needs to be adjusted according to the training results of the model.

In terms of the evaluation metrics, pixel accuracy(PA) and mean intersection over union (mIoU) are the popular metrics used in evaluating the performance of the binary segmentation model.

PA can be represented by Equation (13).

$$PA = \frac{\sum_{i=0}^{K} p_{ii}}{\sum_{i=0}^{K} \sum_{i=0}^{K} p_{ij}} = \frac{TP + TN}{TP + TN + FP + FN}$$
(13)

where *K* represents the foreground classes, and p_{ij} is the number of pixels of class i predicted as belonging to class j. In binary classification problems, they can be simplified as a true prediction fraction, divided by the total number of pixels. where TP and TN represent the true positive fraction and the true negative fraction, and FP, FN represent the false positive fraction and the false negative fraction, respectively.

mIoU is defined as the average IoU over all classes. IoU is defined as the area of intersection between the predicted segmentation result and the ground truth, divided by the area of union between the predicted segmentation result and the ground truth:

$$IoU = \frac{|P \cap G|}{|P \cup G|},\tag{14}$$

where *P* and *G* refer to the predicted result and the ground truth. IoU ranges between 0 and 1.

As mentioned in Section 2.3, the spacecraft segmentation is simplified as a binary segmentation task; thus, mIoU in this paper can be expressed as:

$$mIoU = \frac{1}{2}(IoU_P + IoU_N) = \frac{1}{2}(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FN + FP}),$$
(15)

where IoU_P and IoU_N are the intersection over union of the positive pixels and the negative pixels, respectively.

3.3. Segmentation Performance of SFSNet

In this section, the performance of SFSNet is analyzed and discussed under different experimental parameters and conditions. After SFSNet is initialized and regularized, the augmented image data is continuously generated by the image generator and brought into the model for training.

Figure 6 shows the partial spacecraft segmentation performance of the SFSNet. It can be seen from Table 4 that the loss of SFSNet is as low as 0.062, and the segmentation accuracy reaches 0.965. These data show that SFSNet can precisely segment the spacecraft feature under various scales at current resolution, which solves the issue of spacecraft receptive field variation discussed in Section 1.



Figure 6. Segmentation performance of SFSNet. Image data in the top row are from parts of the test set, and their corresponding segmentation results are laid out in the bottom row.

Furthermore, SFSNet achieves pixel-level segmentation accuracy and has excellent robustness. It can precisely extract the correct edge information from the image regions containing MLI, as shown in Figure 7.



Figure 7. Local details of the segmented image data. SFSNet can accurately extract the edges of the spacecraft and the docking rings from interference caused by the multilayer insulation material.

In terms of SFSNet training procession, we considered the limitations of the device in the field experiment. Due to memory size limitations, the training loss and accuracy of SFSNet under different batchsize condition were compared, as shown in Figure 8.

The results indicated that employing a larger batchsize when training SFSNet led to a decrease in the final loss of the model, while having a minimal impact on its final accuracy. Table 4 also shows that the change in batchsize had little influence on the segmentation performance of SFSNet. The role of batchsize was reflected more in the training process of the model. Due to the limitations of memory and video memory, the training model with



batchsize = 4 is selected as the optimal experimental model in this article. The size of the optimal model is 121.6 Mb.

Figure 8. Training loss and accuracy of the SFSNet under various batchsizes. The abscissa represents the training steps, and the ordinate represents the relative value of the loss and accuracy, respectively.

Table 4. SFSNet performance for different batchsizes.

Batchsize	PA	mIoU
2	0.963	0.927
4	0.965	0.930

As mentioned in Section 3.1, image data captured during spacecraft motion can occasionally exhibit defocus blur due to inherent focus sensitivity constraints. To evaluate the generalization performance of SFSNet in this context, we employed various degrees of defocus blur data for testing, as depicted in Figure 9 and Table 5.

In the data augmentation stage, image data with a defocus radius under 10 pixels are randomly generated by the data generator and brought into the model for training, as described in Table 3. As indicated from Figure 9a–c and Table 5, although the segmentation accuracy decreases slightly with increase in the defocus radius, SFSNet can still correctly segment the target region.



Figure 9. Segmentation performance of defocus blur image of SFSNet. The defocus radius increases gradually from left to right.

With further increase in the defocus radius, Figure 9d and the data presented in Table 5 indicate that several tiny false segmentation areas appeared in the segmentation images, with the PA and mIoU indicators in this group declining accordingly. However, these individual NP samples do not influence the subsequent edge-fitting process as their interference to the expected edge data can be eliminated by the Euclidean distance parameter of the fitting constraints.

Ultimately, when the defocus radius is too large relative to the resolution, the pixellevel information of the spacecraft is almost destroyed, as shown in Figure 9e. In this case, there is too much missing valid input information of the model, which makes it unable to generate a complete segmentation region, and the evaluation indexes drop sharply. In practical experiments, this degree of defocus blur is rarely seen.

Table 5. SFSNet performance under various degrees of defocus blur.

Defocus Radius (px)	0	5	10	15	40
PA	0.965	0.962	0.956	0.944	0.563
mIoU	0.930	0.923	0.912	0.891	0.380

Table 6 shows the SFSNet performances compared to other well-known algorithms. It can be observed that the segmentation accuracy of the FCN network based on VGGNet is lower than that of our method. Although the segmentation results of the Deeplab framework based on ResNet-101 are slightly better than those of SFSNet, its model size reaches 255.4 Mb, which is more than twice the size of SFSNet. The large computational model is not conducive to model pruning and optimization in the subsequent migration process to embedded devices. Therefore, combining segmentation accuracy and model size, SFSNet has an advantage over these mainstream models.

Method	PA	mIoU	Model Size (Mb)
VGGNet	0.830	0.701	103.1
ResNet-101	0.970	0.942	255.4
ouis	0.200	0.930	121.0

Table 6. Comparison of segmentation performances of different models.

In summary, the corresponding segmentation results show that SFSNet has strong robustness and generalization ability during spacecraft motion.

4. Conclusions

This paper has proposed an end-to-end spacecraft image segmentation model based on a convolutional neural network, named SFSNet. SFSNet is not only effective for implementing the precise segmentation of a spacecraft in different imaging scales at a cost of small computation complexity, but it also has strong robustness when processing motion-blurred images. Specifically, the contributions of SFSNet are as follows:

- (1) The segmentation method in this article involves constructing an encoding-decoding structure by a full convolutional network, which replaces the traditional divide and conquer segmentation algorithm. This deep learning model effectively addresses the geometric constraint uncertainty of the spacecraft when it moves within the depth of the field of view, and no additional prior conditions are required. Moreover, compared with the divide and conquer algorithm, SFSNet can effectively suppress the severe influence of the texture of MLI material and segment the spacecraft features with a small error in the ground test.
- (2) A concatenate module is utilized between the encoding and decoding stages of the backbone. The concatenate module combines feature map layers containing pixel-level and higher semantic information, enabling SFSNet to learn finer image details, thus ensuring accurate segmentation accuracy even at high image resolution.
- (3) The multi-objective segmentation task (spacecraft outer contour, spacecraft docking ring, and background) is transformed into a binary segmentation task by leveraging the characteristics of spacecraft motion in the rendezvous stage. This binary task focuses on the internal region of the spacecraft's exterior contour and the external region of the circular berthing ring, simplifying the loss function and the calculation process of SFSNet.

In the experiment, data augmentation was used to enhance the dataset by basic rigidity transformation and defocus blur according to the possible situation of the spacecraft motion

in the ground test. For the high-resolution images in this paper, the capacity of the dataset was able to be greatly expanded by the data augmentation without increasing the workload of manual annotation.

The experimental results show that the proposed method can accurately segment spacecraft features at different scales under the imaging interference effects of multilayer insulation material. During training with batch sizes of two and four, the mIoU of SFSNet reached 0.927 and 0.930, respectively. When training with batchsize = 4, the model exhibited a slower convergence rate compared to batchsize = 2, while maintaining a slightly lower loss. Furthermore, SFSNet trained with the proposed dataset demonstrated competence in segmentation tasks with a defocus radius below 15, achieving a PA of 0.944 and an mIoU of 0.891.

Finally, we compared the segmentation performance of SFSNet with the current mainstream methods. The results showed that, although the model size of SFSNet was slightly larger than that of VGGNet, its segmentation accuracy was much higher than that of the latter. Compared with ResetNet-101, SFSNet's segmentation accuracy was slightly lower, but its model size was only half that of ResetNet-101. These advantages of simple structure and small memory consumption are more conducive to subsequent model pruning and optimization when porting to embedded devices. Therefore, SFSNet is more suitable for ground test applications than the two methods mentioned above.

In future work, we will carry out further research on data expansion and reinforcement learning for specific scenarios to increase the types of spacecraft that can be segmented. In addition, We will also optimize the modular structure of the network in order to simplify the model parameters and improve the training efficiency without losing segmentation accuracy.

Author Contributions: Conceptualization, Y.L. and J.H.; methodology, Y.L.; software, Y.L.; validation, Y.L.; investigation, Y.L.; resources, P.M. and J.H.; data curation, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L.; visualization, Y.L.; supervision, P.M. and J.H.; project administration, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (No. 61473100).

Data Availability Statement: Due to the confidentiality of the research and the requirements of relevant collaborating units, the experimental data involved in this paper are not accessible.

Conflicts of Interest: The authors declare that there are no conflict of interest regarding the publication of this article.

References

- 1. Opromolla, R.; Fasano, G.; Rufino, G.; Grassi, M. A review of cooperative and uncooperative spacecraft pose determination techniques for close-proximity operations. *Prog. Aerosp. Sci.* 2017, *93*, 53–72. [CrossRef]
- Song, J.; Rondao, D.; Aouf, N. Deep learning-based spacecraft relative navigation methods: A survey. Acta Astronaut. 2022, 191, 22–40. [CrossRef]
- Yang, L.; Wang, B.; Zhang, R.; Zhou, H.; Wang, R. Analysis on location accuracy for the binocular stereo vision system. *IEEE Photonics J.* 2017, 10, 1–16. [CrossRef]
- Nedevschi, S.; Danescu, R.; Frentiu, D.; Marita, T.; Oniga, F.; Pocol, C.; Schmidt, R.; Graf, T. High accuracy stereo vision system for far distance obstacle detection. In Proceedings of the IEEE Intelligent Vehicles Symposium, Parma, Italy, 14–17 June 2004; pp. 292–297.
- Zhang, G.; Liu, H.; Wang, J.; Jiang, Z. Vision-based system for satellite on-orbit self-servicing. In Proceedings of the 2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Xi'an, China, 2–5 July 2008; pp. 296–301.
- 6. Mahendrakar, T.; White, R.T.; Wilde, M.; Tiwari, M. Spaceyolo: A human-inspired model for real-time, on-board spacecraft feature detection. In Proceedings of the 2023 IEEE Aerospace Conference, Big Sky, MT, USA, 4–11 March 2023; pp. 1–11.
- Lee, C.; Landgrebe, D.A. Feature extraction based on decision boundaries. *IEEE Trans. Pattern Anal. Mach. Intell.* 1993, 15, 388–400. [CrossRef]
- Brumby, S.P.; Theiler, J.P.; Perkins, S.J.; Harvey, N.R.; Szymanski, J.J.; Bloch, J.J.; Mitchell, M. Investigation of image feature extraction by a genetic algorithm. In Proceedings of the Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation II, Denver, CO, USA, 18–23 July 1999; International Society for Optics and Photonics; Volume 3812, pp. 24–31.

- 9. Zhang, Z. Flexible camera calibration by viewing a plane from unknown orientations. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 1; pp. 666–673.
- 10. Maini, R.; Aggarwal, H. Study and comparison of various image edge detection techniques. *Int. J. Image Process. (IJIP)* **2009**, *3*, 1–11.
- Huo, Y.K.; Wei, G.; Zhang, Y.D.; Wu, L.N. An adaptive threshold for the Canny Operator of edge detection. In Proceedings of the 2010 International Conference on Image Analysis and Signal Processing, Zhejiang, China, 9–11 April 2010; pp. 371–374.
- Miao, X.; Zhu, F.; Hao, Y. Pose estimation of non-cooperative spacecraft based on collaboration of space-ground and rectangle feature. In Proceedings of the International Symposium on Photoelectronic Detection and Imaging 2011: Space Exploration Technologies and Applications, Beijing, China, 24–26 May 2011; Volume 8196, pp. 230–237.
- 13. Peng, J.; Xu, W.; Yan, L.; Pan, E.; Liang, B.; Wu, A.G. A pose measurement method of a space noncooperative target based on maximum outer contour recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2019**, *56*, 512–526. [CrossRef]
- 14. Zhang, F.; Huang, P.; Chen, L.; Cai, J. Line-based simultaneous detection and tracking of triangles. *J. Aerosp. Eng.* 2018, 31, 04018013. [CrossRef]
- Liu, X.; Li, D.; Dong, N.; Ip, W.H.; Yung, K.L. Noncooperative target detection of spacecraft objects based on artificial bee colony algorithm. *IEEE Intell. Syst.* 2019, 34, 3–15. [CrossRef]
- Wu, Y.; Yang, N.; Chen, Z.; Hua, B. Multi-feature fusion based relative pose adaptive estimation for on-orbit servicing of non-cooperative spacecraft. J. Harbin Inst. Technol. (New Ser.) 2019, 26, 19–30.
- Wei, E.; Wei, C.; Zhao, Y.; Jing, Y.; Wang, D. Resident Space Objects Streak Extraction and Angular Measurement Error Analysis Base on Space Image Synthesis System. In Proceedings of the 2018 9th International Conference on Mechanical and Aerospace Engineering (ICMAE), Budapest, Hungary, 10–13 July 2018; pp. 279–283.
- 18. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv* **2014**, arXiv:1409.1259.
- Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
- 20. Ding, L.; Goshtasby, A. On the Canny edge detector. *Pattern Recognit.* 2001, 34, 721–725. [CrossRef]
- 21. Wang, X. Laplacian operator-based edge detectors. IEEE Trans. Pattern Anal. Mach. Intell. 2007, 29, 886–890. [CrossRef] [PubMed]
- 22. Zhang, Q.s.; Zhu, S.C. Visual interpretability for deep learning: A survey. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 27–39. [CrossRef]
- 23. Xu, B.; Wang, N.; Chen, T.; Li, M. Empirical evaluation of rectified activations in convolutional network. *arXiv* 2015, arXiv:1505.00853.
- 24. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
- 25. Zhao, N.; Wu, Z.; Lau, R.W.; Lin, S. What makes instance discrimination good for transfer learning? arXiv 2020, arXiv:2006.06606.
- 26. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. J. Big Data 2019, 6, 1–48. [CrossRef]
- Choi, E.; Biswal, S.; Malin, B.; Duke, J.; Stewart, W.F.; Sun, J. Generating multi-label discrete patient records using generative adversarial networks. In Proceedings of the Machine Learning for Healthcare Conference, Boston, MA, 18–19 August 2017; pp. 286–305.
- 28. Chen, Y.; Zhencong, W.; Wang, M. Design and application of optical system for high-resolution micro space-borne camera. *J. Appl. Opt.* **2020**, *41*, 235–241.
- 29. Yingxiao, L.; Ju, H.; Ping, M.; Jiang, R. Target localization method of non-cooperative spacecraft on on-orbit service. *Chin. J. Aeronaut.* **2022**, *35*, 336–348.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1026–1034.

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