

Article

SDebrisNet: A Spatial–Temporal Saliency Network for Space Debris Detection

Jiang Tao ¹, Yunfeng Cao ^{1,*} and Meng Ding ²

¹ Department of Aerospace Control, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; tjnuaa@nuaa.edu.cn

² College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; nuaa_dm@nuaa.edu.cn

* Correspondence: cyfac@nuaa.edu.cn

Abstract: The rapidly growing number of space activities is generating numerous space debris, which greatly threatens the safety of space operations. Therefore, space-based space debris surveillance is crucial for the early avoidance of spacecraft emergencies. With the progress in computer vision technology, space debris detection using optical sensors has become a promising solution. However, detecting space debris at far ranges is challenging due to its limited imaging size and unknown movement characteristics. In this paper, we propose a space debris saliency detection algorithm called SDebrisNet. The algorithm utilizes a convolutional neural network (CNN) to take into account both spatial and temporal data from sequential video images, which aim to assist in detecting small and moving space debris. Firstly, taking into account the limited resource of the space-based computational platform, a MobileNet-based space debris feature extraction structure was constructed to make the overall model more lightweight. In particular, an enhanced spatial feature module is introduced to strengthen the spatial details of small objects. Secondly, based on attention mechanisms, a constrained self-attention (CSA) module is applied to learn the spatiotemporal data from the sequential images. Finally, a space debris dataset was constructed for algorithm evaluation. The experimental results demonstrate that the method proposed in this paper is robust for detecting moving space debris with a low signal-to-noise ratio in the video. Compared to the NODAMI method, SDebrisNet shows improvements of 3.5% and 1.7% in terms of detection probability and the false alarm rate, respectively.

Keywords: space debris detection; space-based surveillance; lightweight neural network; small object detection; video salient object detection



Citation: Tao, J.; Cao, Y.; Ding, M. SDebrisNet: A Spatial–Temporal Saliency Network for Space Debris Detection. *Appl. Sci.* **2023**, *13*, 4955. <https://doi.org/10.3390/app13084955>

Academic Editor: Sungho Kim

Received: 10 February 2023

Revised: 8 March 2023

Accepted: 10 March 2023

Published: 14 April 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/>).

1. Introduction

With the advancements in space technology, the increasing number of operational spacecraft is facing a severe threat from space debris, which is caused by frequent space launch activities [1,2]. According to the U.S. As of March 2022, the Space Surveillance Network (SSN) has cataloged up to 25,000 space debris, defunct spacecraft, and active spacecraft, with numbers expected to increase continuously in the future [3]. Collisions with large space debris could destroy an entire spacecraft, while small-sized space debris could cause irreversible damage to spacecraft due to their high velocities, such as performance degradation or dysfunction. Therefore, it is vital to observe space debris in the distance to avoid latent collisions with spacecraft. One crucial step involves space debris detection, which could be further used for space debris tracking and cataloging.

There are currently two main categories of space debris surveillance systems: space-based systems and ground-based systems. Ground-based systems rely on large telescopes or radar installed on the ground for space debris detection and recognition. In contrast, space-based systems use onboard sensing devices to detect space debris. The advantages of space-based systems over ground-based systems can be described as follows. (1) They

are not affected by weather or the circadian rhythm. (2) They could avoid the limits of stationary observation sites. (3) They could detect millimeter-sized small objects; the latter is aimed at centimeter-sized objects [4]. Therefore, space-based surveillance systems are effective measures for enhancing the safety levels of spacecraft. For the application of space debris surveillance, onboard sensing devices generally contain visible sensors, infrared sensors, and radars. Among the sensing devices, visible sensors are promising solutions due to their superior autonomous levels and highly accurate observation data [5].

The typical procedure for detecting space debris involves object extraction and centroid computation [6,7]. The centroid can then be used for astrometry by matching the observed star field with the star catalog [8]. However, the main focus of this paper is object extraction; a simple centroid computational method was adopted to evaluate the proposed object extraction method. This paper does not cover the astrometry problem. Based on different operational modes of the surveillance system, commonly used visible sensor-based space debris detection strategies can be divided into streak-like object detection and point-like object detection. When the orientation of the surveillance platform is continuously fixed to stars, the space objects appear as streak-like regions; this operational mode is called sidereal tracking. When the surveillance platform is continuously reoriented to be fixed at the space objects, the objects appear as point-like regions; this mode is called object tracking [9]. Afterward, space debris and stars could be distinguished by their shape characteristics in both operation modes using only single-frame images. Typical single-frame image-based methods for space debris detection include the Hough transform method [10], feature-based methods using image moments [7], point spread function (PSF) fitting techniques [11,12], mathematical morphological methods [13,14], etc. However, faint space debris could not be detected by single-frame images due to the tiny observable magnitude limited by charge-coupled device (CCD) sensors [15]. To deal with this problem, multi-frame image-based methods have been proposed using the space object's motion information, which can be divided into two categories: track-before-detect (TBD) methods and optical flow methods. The main idea of TBD is to improve the signal-to-noise ratio (SNR) of faint space objects according to accumulating measurements to produce more confident detections. The representative TBD-based space debris detection methods include the stacking method [16,17], line-identifying technique [18], particle filter (PF) [19,20], and multistage hypothesis testing (MHT)-based methods [21–23]. However, there are many limitations to these methods. The stacking method needs to presume a number of likely movement paths and requires time to analyze the observation data for the uncatalogued space debris. The line-identifying technique assumes the space debris moving with a constant velocity and the analysis time will rise exponentially with the number of candidate detections in each frame. PF-based methods need to carefully design the likelihood function to confirm the trade-off between particle convergence and final tracks under the computational restrictions. For MHT-based methods, candidate trajectories will increase rapidly with the number of space objects and noise, resulting in high computational costs. The optical flow-based methods [24–26] can be performed without prior information but they are also limited by their high computational complexity.

With the progress in satellite imaging technology, video satellite image data have been introduced to detect space objects in recent years [27,28]. Different from multi-frame image sequences taken with long exposure times and long interval times, video image data can capture spatial information as well as more compact temporal information with the continuous movement state of space objects, which is vital for the immediate orbital anomaly detection and continuous monitoring. Typical video satellites and their key parameters are listed in Table 1.

Table 1. Typical video satellites and key parameters.

| Satellite Name | Field of View (°) | Resolution (Pixel × Pixel) | Frame Rate (Frames/s) |
|-----------------|-------------------|----------------------------|-----------------------|
| UrtheCast [29] | - | 1920 × 1080 | 3 |
| TUBSAT [30] | 0.28 × 0.21 | 752 × 582 | - |
| SkySat [31] | - | 1920 × 1080 | 30 |
| Tiantuo-2 [28] | 2.5 × 2.5 | 960 × 576 | 25 |
| Jilin-1 [29,32] | - | 3840 × 2160 | 25 |

Zhang et al. [28] proposed a space object detection algorithm based on motion information using the Tiantuo-2 video satellite. The author removed the image background based on local image properties for a single-frame first. The space object trajectories are associated with the Kalman filter for all frames. Finally, the trajectory is considered a space object if the mean velocity of the object in all frames is larger than a given threshold. The algorithm can detect moving space objects with brightness changes from satellite videos, but it is not a straightforward process. Moreover, many parameters should be set up through empiricism or experiment in advance. The advantages and disadvantages of the above three methods are summarized in Table 2.

Table 2. The advantages and disadvantages of different space debris detection methods.

| Methods | Advantages | Disadvantages |
|---------------------------|--|--|
| Single-frame-based method | It is straightforward. | (1) It needs a priori information and many predefined templates. (2) The faint and moving objects could not be detected. |
| Multi-frame-based method | The faint and moving objects could be detected. | (1) It is time-consuming. (2) It needs a priori information. |
| Video-based method | (1) The faint and moving objects could be detected. (2) It does not need a priori information. | The video data have low signal-to-noise ratios due to the short exposure times. |

Over the past decade, deep learning has made great advancements in computer vision [33], and deep convolutional neural networks have shown significant competence for video salient object detection (VSOD) [34–36]. The central principle of the VSOD involves learning temporal dynamic cues related to moving objects. based on the fact that long-range dependencies exist in the space and time of consecutive frames [37,38]. Yan [39] proposed a non-locally enhanced temporal module to construct the spatiotemporal connection between the features of input video frames. Chen [34] developed a nonlocal self-attention scheme to capture the global information in the video frame. The intra-frame contrastive loss helps separate the foreground and background features and inter-frame contrastive loss improves temporal consistency. Su [37] introduced the transformer block to capture the long-range dependencies among group-based images through the self-attention mechanism and designed an intra-MLP learning module to avoid partial activation to further enhance the network. However, using VSOD directly for space debris detection is questionable. The reasons can be concluded as follows: (1) The large distances involved in detecting space debris can make the images appear very small, and they are often referred to as “small objects”. Therefore, it is difficult to extract spatial features of space debris in videos by deep neural networks. (2) The cluttered backgrounds generated by large numbers of stars appearing as point source objects can make it difficult to distinguish space debris from star backgrounds. (3) Various noises exist in space surveillance platforms, such as thermal noise, shot noise, dark current noise, and stray light. Therefore, space debris energy is very faint compared to the noise background, which also brings challenges to space debris detection.

To solve these difficulties, we propose a novel spatial–temporal saliency method called SDebrisNet for detecting space debris in satellite videos; it uses lightweight atrous spatial

pyramid pooling and scale-enhanced structures to enhance context-aware and scale-aware features, while constrained self-attention helps capture local temporal features. First of all, a spatial feature extraction module (SFM) is proposed to improve the feature extraction performance for small objects. We added a high-resolution feature detector to the multi-scale feature extraction network to extract more specific features from the low-level feature map. Then, the spatial data of small objects were further strengthened by a spatial feature enhancement module (SFEM). Next, the spatial–temporal coherence was enhanced by a temporal feature extraction module (TFM). Finally, the saliency map of space debris can be obtained by a spatial–temporal feature fusion module (STFM). The main contributions of this paper can be summarized as follows.

(1) In order to account for the unique characteristics of moving space debris in satellite videos, a spatial–temporal saliency framework was developed. This framework enables end-to-end space debris detection without the need for preprocessing or post-processing steps.

(2) SFM extracts the spatial features using a lightweight neural network, which is well-suited for space-based computational platforms with limited storage and computational resources.

(3) To deal with the small object detection problem, SFEM enhances both the context-aware features and the scale-aware features, which could significantly improve object detection precision in different scales, including small objects.

(4) TFM connects both the spatial features and temporal features from the consecutive frames. Even if space debris data possess a low signal-to-noise ratio, our method could effectively output the saliency maps. The novelty of this paper can be summarized as follows: (1) A new space debris detection method based on spatial–temporal saliency is proposed. (2) A new saliency detection neural network based on the constrained self-attention is proposed. (3) A new small object feature extraction network, including a spatial feature extraction network and spatial feature enhancement network, is proposed. The rest of the paper is organized as follows. Section 2 describes the proposed spatial–temporal saliency network architecture for space debris detection. The centroid computational method is introduced in Section 3. Section 4 presents the experimental setup. Section 5 presents experimental studies for verifying the proposed space debris detection method, including the comparison results with the current space object detection algorithms. Finally, we conclude our research in Section 6.

2. Space Debris Detection

2.1. Overview of the Proposed Detection Method

Figure 1 shows the flow chart of the proposed space debris detection method, which mainly includes saliency detection and centroid computation. Saliency detection uses video clips as input and output corresponding saliency maps, from which centroid coordinates of space debris for each frame can be computed.

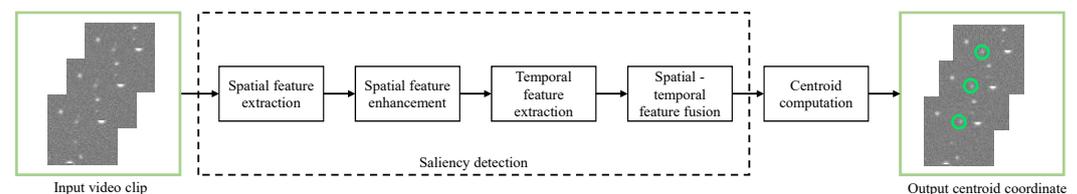


Figure 1. Flow chart of the proposed space debris detection method.

Figure 2 shows the overall structure of the proposed spatial–temporal method for space debris detection in satellite videos, which contains four major modules: SFM, SFEM, TFM and STFM. First, the video clip contains several consecutive frames $\{I_i\}_{i=1}^N$, where N is the number of frames, which are fed into the SFM, which is aimed at extracting the spatial features $\{C_i\}_{i=1}^N$ from the raw input frames and outputting five different resolutions of feature maps $\{C^j\}_{j=1}^5$ for each frame. Second, SFEM consists of four spatial feature

enhancement (SFE) networks designed to enhance the spatial features outputted by SFM and the five enhanced features $\{CE^j\}_{j=1}^5$ with the same shape of $\{C^j\}_{j=1}^5$. Meanwhile, each high-level feature map C_i^5 of each frame is combined as $C_S \in \mathbb{R}^{C \times N \times H \times W}$ for a video clip by a concatenation operation, where C, H, W denote the channels, height, and width of the feature maps, respectively. Then, C_S is sent to the TFM to learn the spatiotemporal coherence of the consecutive frames based on the self-attention method and outputs the temporal salient high-level features C_T . Finally, the resulting saliency images $\{S_i\}_{i=1}^N$ with values in the range $[0, 1]$ of the video clip, were obtained through STFMs, composed of the residual connected refinement network, by fusing the enhanced spatial features and temporal features from high to low levels, progressively.

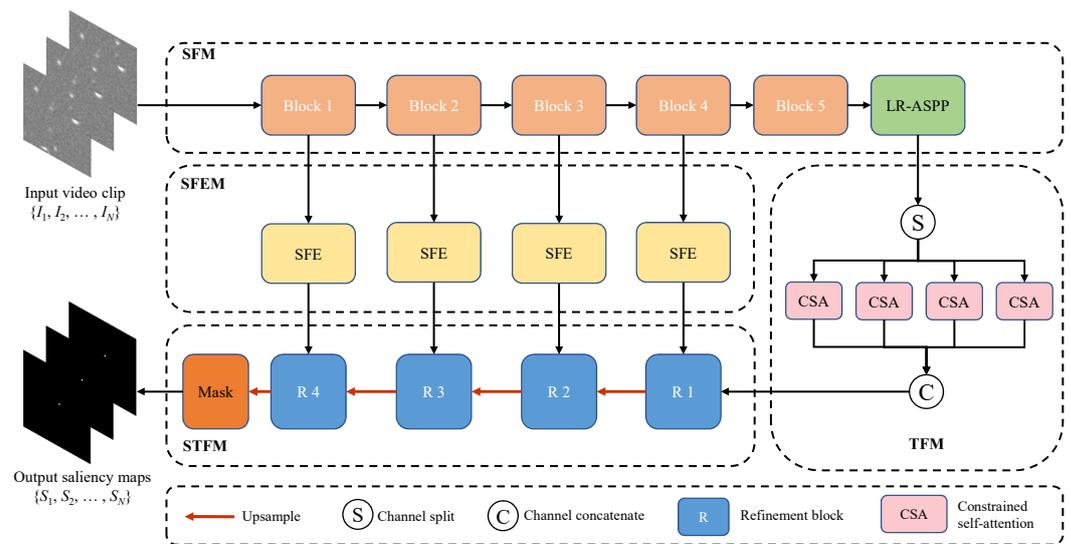


Figure 2. Detailed illustration of the proposed saliency detection network, which consists of the spatial feature extraction module, spatial feature enhancement module, temporal feature extraction module, and saliency prediction module.

2.2. Spatial Feature Extraction

Many high-performance feature extractors, such as ResNet [40], Xception [41], etc., have been proposed in recent years. However, these models possess massive model parameters (typically tens of millions) and high floating-point operations (FLOPs), which are not well-suited to be deployed on current space-based platforms with limited memory and computational resources [42]. Accordingly, MobileNetV3 [43], one of the most popular lightweight backbones, is modified as the spatial features extraction module. MobileNetV3 has fewer parameters, lower FLOPs, and faster inference speeds. Moreover, it is well-matched for low resource-use cases and achieves a high accuracy-latency trade-off with mobile devices. The proposed spatial feature extraction network is shown in Figure 3. It mainly consists of a convolutional layer with a kernel size of 3×3 and a series of bottleneck structures [43]. The bottleneck structure is a resource-efficient block composed of an inverted residual structure and linear bottleneck layers with squeeze and excitation modules [44]. The inverted residual structure could improve the ability of a gradient to propagate across multiplier layers as well as allow for considerable memory-efficient implementation. The linear bottleneck layers could prevent the information loss caused by non-linear transformations, such as rectified linear unit (ReLU) activation. The squeeze and excitation module is a channel-wise attention module applied in the linear bottleneck layers to focus on the most important feature representations.

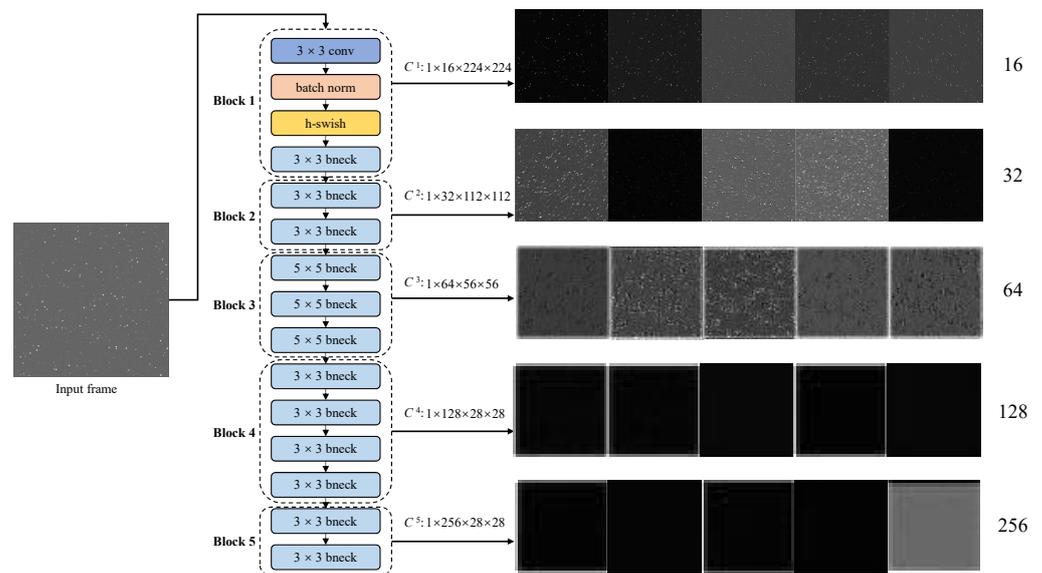


Figure 3. Spatial feature extraction module. The black squares on each line denote the first five outputted feature maps of each block. The number at the end of each line means the number of extracted feature maps.

Compared to the original MobileNetV3 model, three modifications were performed to make it applicable for space debris feature extraction.

1. Some layers of the original model are tailored to further reduce the model parameters to adapt to space-based applications. Specifically, the last three bottleneck blocks were removed. This is because the space debris in the satellite video frame had a simple structure and small size, which did not need a deeper network to extract complex semantic features or a large receptive field to detect the small objects.

2. The SFM was divided into five blocks and five different resolutions of feature maps were outputted correspondingly. The feature maps from shallow layers with higher resolutions preserved detailed features, such as position and intensity, which are beneficial for small object feature extraction. The high-level features with large receptive fields contain rich semantic features such as contours, which are important for learning temporal cues and distinguishing salient objects [45]. Based on the above facts, we used the first four blocks to extract multi-scale spatial features and the last block to learn temporal features.

3. A lightweight atrous spatial pyramid pooling (LR-ASPP) [43] module was added to the last block, which could capture the image-level global context features and improve the inference speed.

Figure 3 shows the first five outputted feature maps of each block, which have half the input resolution of the previous block. We can see that the first two blocks respond to detailed corners and color conjunctions. Block 3 only outputs more coarse color and position information. The last two blocks almost don't extract any semantic features. Therefore, high-level features should be strengthened for spatial-temporal feature learning.

2.3. Spatial Feature Enhancement

As contextual information is crucial for finding small objects [46], the context-aware features were obtained first. The light-weight atrous spatial pyramid pooling [43] module (LR-ASPP), with a large pooling kernel and a 1×1 convolution, was employed to generate context features. Thus, four LR-ASPP modules were connected to the first four blocks of the

backbone, respectively, to capture the image-level global context. Inspired by the scale enhancement structure in [47], we delivered the outputted context-aware features to the scale enhancement module to further strengthen the scale cues of small objects. The generated scale-aware features F_k^o of the given context-aware features F_k^i can be computed as:

$$F_k^o = (1 + F_k^i) \odot C_k \tag{1}$$

where C_k are the feature maps outputted by SFM at different blocks, as shown in Figure 3. ‘ \odot ’ and ‘+’ refer to element-wise multiplication and element-wise addition, respectively. Since the spatial features are also vital to saliency detection, the skip connection structure was adopted to mitigate the feature degradation around the objects caused by model singularities. Lastly, we attached the residual skip connection layer [40] to each SFEM to enable the connection between the extracted spatial features and the saliency prediction model. The residual skip connection layer here is different from the above skip connection structure, which only skips the element-wise multiplication operation. The residual skip connection layer consists of three convolutional layers, which could downsample the channels of feature maps at different layers to the same channels as well as share powerful target information with the spatial-temporal feature fusion module.

Figure 4 shows the enhanced spatial features of space debris. Compared to the spatial features generated by SFM, more spatial features are obtained through SFEM. For example, many colors and corners are revealed in column 2 and column 5 of block 2 in Figure 4, while no features are extracted by SFM in Figure 3. The higher layers can extract high-level features, such as the shapes and contours of objects (e.g., row 4 in Figure 4) and more complex textures (e.g., row 5 in Figure 4), while the SFM cannot, as shown in Figure 3.

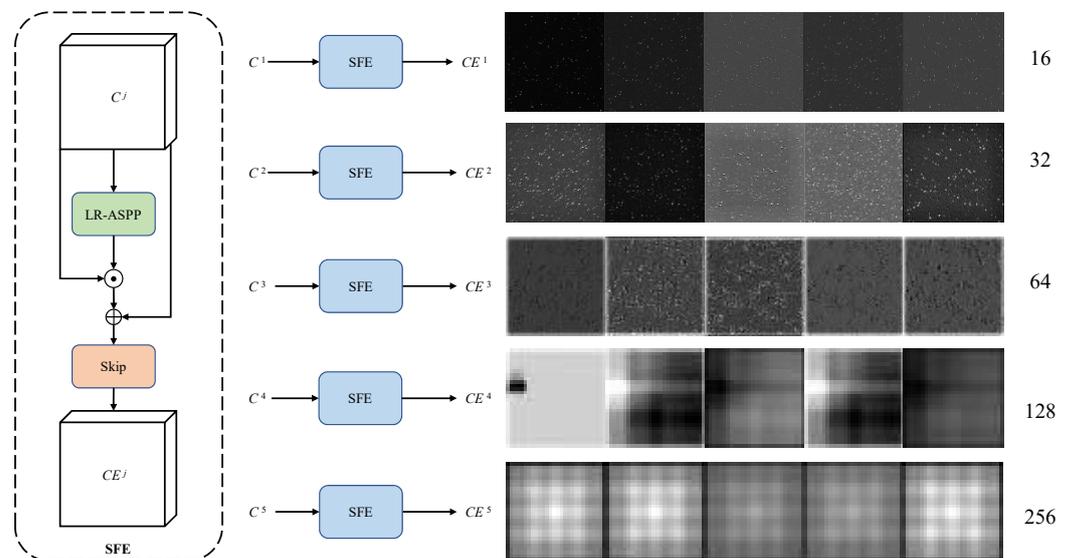


Figure 4. Spatial feature enhancement module.

2.4. Temporal Feature Extraction

The mainstream temporal feature extraction approaches mainly include 3D-convolution [48], recurrent neural network [39,49], optical flow [35,50] and attention-based mechanisms [34,37,51]. However, the computational cost of the 3D convolution operation is usually expensive, which is not fit for the space-based platform with limited computational resources. For the recurrent network, the convolutional memory units, such as convolutional long short-term memory (ConvLSTM) and convolutional gated recurrent unit (ConvGRU) lead to high computational costs and high memory costs. The optical flow method requires an additional network branch to obtain optical flow information, which means it is not a real end-to-end network. Moreover, when the camera moves due to the incident jitter of the satellite video, the movement of space debris may be too small

compared to the background, resulting in a weak correlation between salient objects and movement information contained in optical flow. Recently, non-local self-attention-based methods [34,39,52] have been widely applied to the video salient object detection to model the temporal features. They can capture long-range dependencies in videos by establishing pair-wise relationships among feature elements in consecutive frames. However, they cannot be applied in space debris detection in videos (directly) for the following reasons. Firstly, they focus on capturing motion-independent global contexts instead of motion cues, which are not suitable for detecting moving space debris because the moving objects tend to be salient and are more attractive to human attention [53]. Secondly, the computational and memory costs are expensive because the FLOPs (floating point operations) and memory consumption are quadratic functions with respect to the number of frames and the feature resolution, which are not fit for the limited computations and storage resources of space-based applications. Motivated by the constrained self-attention (CSA) operation in [51], which can model motion cues more efficiently, a parallel structure composed of four CSA operations was designed as the TFM to extract the temporal features.

Given the high-level spatial features $C_S \in \mathbb{R}^{C \times N \times H \times W}$ generated by SFM, they are split into four groups, i.e., $C_{S'} \in \mathbb{R}^{\frac{C}{4} \times N \times H \times W}$. Each $C_{S'}$ is fed into a CSA network with different dilations and window sizes. The four resulting features $C_{T'}$ are concatenated as temporal features C_T . The structure of the CSA network can be found in Figure 5. The $C_{S'}$ is projected into three subspaces, i.e., query space \mathbf{Q} , key space \mathbf{K} , and value space \mathbf{V} , using $1 \times 1 \times 1$ convolution as the linear function. Based on the objects sharing similar positions in adjacent frames, when a feature element $q \in \mathbb{R}^{\frac{C}{4} \times 1 \times 1 \times 1}$ is queried in the position (n, h, w) of \mathbf{Q} , where $1 \leq n \leq N, 1 \leq h \leq H, 1 \leq w \leq W$, the constrained neighborhood S_q surrounding q in \mathbf{K} is selected to compute affinity by the dot-product operation. S_q can be formulated as

$$S_q = \{K(n', h', w')\}_{n'=1, h'=h-dr, w'=w-dr}^{N, h+dr, w+dr} \tag{2}$$

where $S_q \in \mathbb{R}^{N(2r+1)^2 \times \frac{C}{4}}$. $K(n', h', w') \in \mathbb{R}^{\frac{C}{4} \times 1 \times 1 \times 1}$, $1 \leq n' \leq N, 1 \leq h' \leq H, 1 \leq w' \leq W$. r and d are the size and dilation of the sliding window, respectively. Thus, the size of the constrained neighborhood can be determined by different sliding window sizes r , dilation d , and the number of video frames N . The affinity function f between feature elements q and S_q can be computed by

$$w_q = f(q, S_q) = \sum_{n'=1, h'=h-dr, w'=w-dr}^{N, h+dr, w+dr} qK(n', h', w')^T \tag{3}$$

where $w_q \in \mathbb{R}^{N(2r+1)^2 \times 1 \times 1 \times 1}$. Then the augmented feature q' in $C_{T'}$ can be calculated as a weighted sum of the embedding feature $V(n'', h'', w'') \in \mathbb{R}^{N(2r+1)^2 \times 1 \times 1 \times 1}$ with weight w_q :

$$q' = w_q V(n'', h'', w'') \tag{4}$$

where $q' \in \mathbb{R}^{\frac{C}{4} \times 1 \times 1 \times 1}$, $1 \leq n'' \leq N, 1 \leq h'' \leq H, 1 \leq w'' \leq W$.

The parallel CSA structure with different window sizes and dilations can capture the multiple scales and various speed cues simultaneously. Obviously, the sliding window size should be larger than the space debris size (but not too large to reduce computational costs). With regard to the speed, a higher value means the object will cross a wider space in the same interval as shown in Figure 6. Based on the space debris size analyzed in Section 4.1, both the window size r and dilation are set to $\{1, 2\}$.

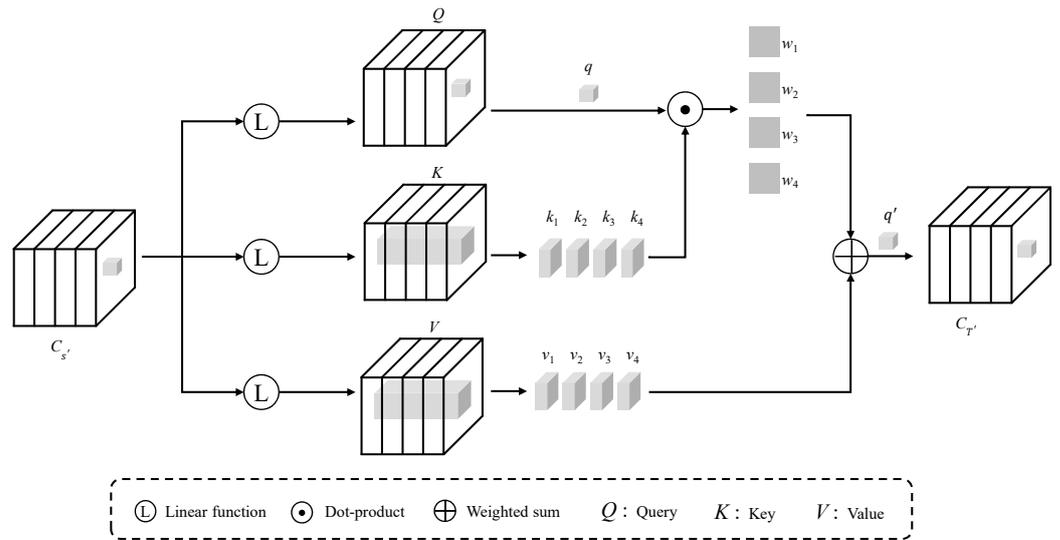


Figure 5. Illustration of the constrained self-attention (CSA) network. The video clip includes four frames as examples in this figure.

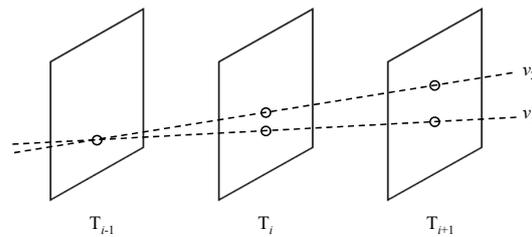


Figure 6. The space positions of space debris at different motion speeds. The space debris cross three consecutive frames at two different speeds, where $v_2 > v_1$.

2.5. Spatial–Temporal Feature Fusion

The STFM is used for outputting the resulting saliency map by fusing the spatial features and temporal features obtained by SFEM and TFM, respectively. The STFM consists of four stacked refinement blocks, each of which takes as input the output feature map F_{td}^i of the previous refinement block in the top-down stream and the output feature map F_{bu}^i from SFEM in the bottom-up stream. This design can mitigate the effects of the loss of spatial details caused by a series of convolutional layers and downsampling operations from SFM. The main workflow of STFM includes three steps: (1) concatenating F_{td}^i and F_{bu}^i , (2) feeding them to a convolutional layer with a 3×3 kernel size and 16 channels, (3) up-sampling the output to ensure that F_{td}^i and F_{bu}^i have the same spatial resolution by a bilinear interpolation operation and that F_{td}^{i+1} can be obtained. Note that F_{td}^1 is the same as the output feature C_T from TFM. Lastly, a probability map with the same resolution as the input image by a decoder composed of two convolutional layers can be generated.

3. Centroid Computation

Once the salient region is acquired, the centroid of space debris in the CCD focal plane can be obtained by a simple energy-weighted average method. The overall aim of space debris detection is to compute the pixel coordinate of the centroid of space debris, which can be further used to evaluate the performance of the proposed object region extraction method. The principle of centroid computation is to assign different weights to the point-source pixels of the extracted salient region depending on the intensity. The weighted centroid algorithm is summarized as follows:

$$x_0 = \frac{\sum_{x=m_1, y=n_1}^{m_2, n_2} xI(x, y)}{\sum_{x=m_1, y=n_1}^{m_2, n_2} I(x, y)} \quad (5)$$

$$y_0 = \frac{\sum_{x=m_1, y=n_1}^{m_2, n_2} yI(x, y)}{\sum_{x=m_1, y=n_1}^{m_2, n_2} I(x, y)} \quad (6)$$

where (x_0, y_0) is the weighted centroid coordinate of the pixel location (x, y) . x and y are the pixel locations of the salient region. $I(x, y)$ is the pixel intensity value at the pixel location (x, y) . m_1, m_2, n_1, n_2 denote the pixel ranges of the salient region. The accuracy of the centroid computational method is sensitive to background noise [7]. However, this effect can be neglected because the saliency detection method outputs a segmentation mask from the background without any noise.

4. Experimental Setup

4.1. SDD Dataset

There is no publicly available space debris dataset for the use of space debris detection in videos at present. Therefore, we created a space debris video saliency detection dataset named SDD (<https://github.com/taojianggit/SDD> (accessed on 27 July 2022)), which includes 45 synthetic video sequences and 2 real video sequences, to assess the quantitative and qualitative performances of the space debris detection algorithm. Each synthetic sequence contains 250 images and 2 real video sequences contain 1551 images. Finally, SDD includes 47 video sequences; 33 synthetic sequences are used for training and the remaining 14 sequences are used for testing. All synthetic video sequences are created by using the same video satellite sensor properties of Tiantuo-2 [28]. The specific parameters of video satellite sensors are provided in Table 3.

Table 3. Video satellite sensor parameters.

| Parameter | Value |
|------------------|-------------------------|
| Resolution | 960 pixels × 576 pixels |
| Focal length | 1000 |
| Field of view | 2.5 × 2.5 |
| Pixel dimensions | 8.33 × 8.33 |
| Frame rate | 25 frame/s |

The SDD dataset covers different SNRs, space debris diameters, space debris motion directions, and motion speeds. Here, the SNR value is defined as $SNR = \frac{S-B}{\sigma}$, where S is the mean pixel value of the object region, B is the mean pixel value of the background region, and σ is the standard deviation of the background region. In general, the background region is three times the size of the object region. We simulated both the high SNR space debris and low SNR space debris with SNR = 10, 5, 1, 0.8, 0.6, 0.4, 0.2, and 0.1. The space debris with diameters = 0.01 m, 0.05 m, 0.1 m, and 0.5 m were simulated. The motion directions of space debris were changed with an interval of 20° from 0° to 340°. The space debris with motion speeds (pixel/frame) = 0.16, 0.20, 0.27, 0.40, 0.44, 1.00, 1.33, 2.00, and 4.00 are defined in the dataset. The examples of synthetic video sequence images of space debris are shown in Figure 7.

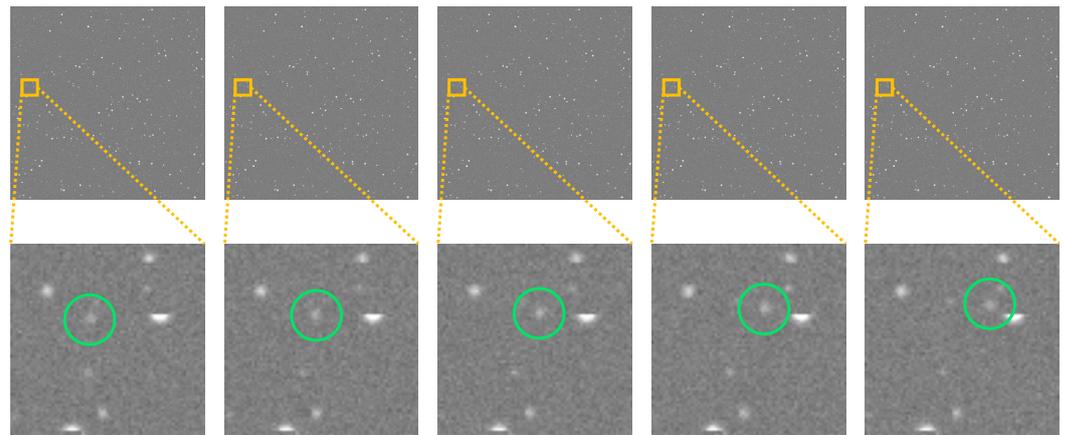


Figure 7. Example of a video sequence with SNR = 0.2, diameter = 0.1 m, speed = 4 pixel/frame, direction = 10° in SDD dataset (top row). Close-up of space debris marked by a green circle in the image sequence (bottom row).

4.2. Metrics

We adopted three groups of metrics to evaluate the proposed method. The maximum F-measure (F_β) [54] and mean absolute error (MAE) were used for evaluating the proposed saliency detection method, as seen in Section 2. The centroid error (CE) was adopted to evaluate the pixel error of the computed centroid coordination based on the saliency detection results. The detection probability P_d and false alarm rate P_f were introduced to compare them with different space debris detection methods. The F-measure is defined as $F_\beta = \frac{(1+\beta^2) \times \text{precision} \times \text{recall}}{\beta^2 \times \text{precision} + \text{recall}}$, which denotes a harmonic average of precision and recall. Precision and recall were obtained by comparing the predicted saliency maps (binarized at every integer threshold in the range of [0, 255]) with their ground truth. β^2 is empirically set to 0.3 as suggested in [55]. The mean absolute error is computed by $MAE = \frac{1}{N} \sum_{k=1}^N |s_k - g_k|$, where $s_k \in S$ and $g_k \in G$ denote the predicted saliency maps and the corresponding ground truth, respectively. k and N refer to the pixel position and pixel number of a saliency map. MAE represents the averaging pixel-wise error between the predicted saliency map and ground truth. S-measure [56] is another popular saliency detection metric that reflects the structural similarity between the predicted salient objects and the ground truth. However, since space debris appears as blobs without distinct structures, S-measure is not utilized in this work. The centroid error is calculated by comparing the ground truth and the computed pixel coordinate obtained by the centroid algorithm. The detection probability P_d and false alarm rate P_f are defined by $P_d = \frac{N_d}{N_{all}}$ and $P_f = \frac{N_f}{N_{all} + N_f}$, respectively. N_d represents the number of correctly identified space debris. The object would be considered as space debris when the computed centroid pixel coordinate falls within the radius of 5 pixels from the center coordinate of ground truth. N_f represents the number of background stars and noise that are incorrectly identified as space debris, while N_{all} represents the total number of space debris. A large assessment score indicates good performance for F_β and P_d , and the reverse for MAE, CE, and P_f .

4.3. Methods for Comparison

The proposed method is compared with three state-of-the-art space object detection methods, which are the channel and space attention U-net (CSAU-Net) [57], the topological sweep (TS) method [58], and the new object detection algorithm using motion information (NODAMI) method [28]. CSAU-Net is a single-frame-based space object detection method that adds attention modules in the traditional encoder and decoder structure to enhance features and better use original feature layers. TS is a multi-frame space object detection method that exploits the geometric duality to find GEO objects from short sequences of optical images. NODAMI is a video-based space object detection method using motion

information from video satellites. CSAU-Net is a recent deep learning-based method aiming to achieve state-of-the-art performance. The related parameters are set to their default values as described in the original publications. With regard to the deep learning-based methods, the pretrained models provided by the corresponding authors were adopted.

4.4. Training Setup

We implemented the proposed SDebrisNet in PyTorch 1.10. The network is trained on the Debian 10 system with a single A100 40 GB GPU, Intel Cascade Lake CPU. The network is initialized by PyTorch default setting and is not pretrained with any dataset. During the training, the batch size was set as 1. We adopted the Adam optimizer with a learning rate of 1×10^{-4} . The input images or video frames were resized to 448×448 before being fed into the network in both the training and inference phases. The sigmoid cross-entropy loss was used as the loss function to compute the loss between each input video frame and the corresponding ground truth saliency map. Moreover, the video clip length was set to 4 in both the training and test stages.

5. Experimental Results and Analysis

This section describes the experimental results of the SDD dataset. Section 5.1 describes the quantitative comparison results with other space debris detection methods. Section 5.2 analyzes the sensitivity of the proposed method to the video clip length. Section 5.3 describes the ablation study. The robustness of the proposed method is assessed in Section 5.4. Section 5.5 provides the test results on the real space object video sequences. All of the test experiments were performed using a computer equipped with an Intel Xeon W-2123 CPU and NVIDIA RTX2080TI GPU.

5.1. Comparison with Other Space Object Detection Methods

We compare the proposed SDebrisNet with some representative space object detection methods, as mentioned in Section 4. However, they are not saliency-based object detection methods, so the saliency detection metrics cannot be used for comparing these methods. The detection probability P_d and false alarm rate P_f as described in Section 4.2 are introduced to evaluate the different space debris detection methods. Table 4 presents the detection results of all the space debris detection methods in the proposed SDD dataset. In Table 4, the bold denotes the best result under the current metric. It can be seen from Table 4 that the multi-frame-based method outperforms the single-frame-based method because it can utilize the interframe motion information. The video-based methods achieve better detection results than both the single-frame-based and multi-frame-based methods. This is because satellite video images have strong interframe correlations and their motionless backgrounds have great redundancy under the sidereal tracking mode. However, the multi-frame satellite images have weak correlations due to the long exposure times between two adjacent frames. As a whole, the proposed SDebrisNet achieves the best performance. Compared with the NODAMI method, SDebrisNet improves by 3.5% and 1.7% in terms of P_d and P_f , respectively.

The computational efficiencies of different space debris detection methods over the image size of 960×576 pixels are also listed in Table 4. On average, our method requires a processing time of 0.013 seconds/frame, which is less than the other three methods. The proposed method could process a frame for the video satellite with a frame rate of 25 frame/s.

5.2. Sensitivity Analysis to the Video Clip T

The video clip length T is a hyperparameter that affects the space debris detection performance of SDebrisNet. It denotes the number of video frames that are fed into the spatial-temporal saliency network for each batch during the training stage. First, a subset of the training set is divided and then it is trained without the pretrained model for each video clip length. The performance of SDebrisNet trained with different T is shown in Table 5. As

shown in the table, SDebrisNet achieves better detection performance from the increase in the video clip length. However, when the value of T is too large, the detection performance will decrease. This is because SDebrisNet cannot sufficiently learn the spatial–temporal information when the value of T is too small. However, a large value of T may cause objects to disappear from the reference window, which can be detrimental to the PCSA’s ability to learn motion cues. In addition, the computational cost will increase with longer video clip lengths. $T = 4$ is the most suitable choice for spatial-temporal information propagation.

Table 4. Detection results of different space debris detection methods on the SDD dataset.

| Method | $P_d \uparrow$ | $P_f \downarrow$ | Running Time |
|---------------------------------|----------------|------------------|----------------|
| <i>singleframe-based method</i> | | | |
| CSAU-Net [57] | 0.912 | 0.021 | 0.021 s |
| <i>multiframe-based method</i> | | | |
| TS [58] | 0.933 | 0.035 | 0.074 s |
| <i>video-based method</i> | | | |
| NODAMI [28] | 0.961 | 0.028 | 0.195 s |
| SDebrisNet | 0.996 | 0.011 | 0.013 s |

Table 5. The detection performance of SDebrisNet for different hyperparameters T .

| T | Precision \uparrow | Recall \uparrow | $F_\beta \uparrow$ | MAE \downarrow |
|-----|----------------------|-------------------|--------------------|------------------|
| 2 | 0.733 | 0.805 | 0.982 | 0.014 |
| 4 | 0.802 | 0.796 | 0.984 | 0.014 |
| 6 | 0.771 | 0.821 | 0.963 | 0.015 |
| 8 | 0.707 | 0.798 | 0.976 | 0.016 |
| 10 | 0.716 | 0.788 | 0.920 | 0.016 |
| 15 | 0.754 | 0.750 | 0.920 | 0.017 |
| 20 | 0.733 | 0.644 | 0.821 | 0.025 |

5.3. Ablation Study

To investigate the effectiveness of SFM, SFEM, and TFM modules in SDebrisNet, the ablation experiment is conducted in this section. To analyze the contributions of individual modules, the saliency detection results of SDebrisNet without different components are shown in Figure 8 and the corresponding quantitative results are recorded in Table 6. As shown in Figure 8, ground truth displays the clearest and strongest streak among all of the detection results. SFM reveals a sparse streak; this is because SFM cannot extract sufficient spatial features due to the tailored feature extraction network. SFEM performs better than SFM due to the more enriched spatial features learned by SFEM. Since the temporal feature extraction module is added on the basis of the original feature extraction network, TFM also performs better than SFM. The performances of (f) and (g) are slightly better than (d) but significantly better than (e). It is because the spatial features and the temporal features are captured by the network. Moreover, the SDebrisNet performs better than (f) and (g) and the best in the ablation study. The reason is that the SFM+SFEM extract the strong spatial features while the TFM learns the motion cues simultaneously. In Table 6, the first row represents the detection results of the baseline composed of original MobileNetV3+SPM, which achieves poor detection results in terms of both F_β and CE due to insufficient spatial–temporal features extracted from the baseline network. The performance is degraded when MobileNetV3 is replaced by SFM due to the tailoring of the original network structure. After adding SFEM, it can create a certain level of performance improvement, which is consistent with the results shown in Figure 8. On the basis of SFM or SFEM, adding the TFM can significantly increase by 15.5% and 16.6% w.r.t F_β , respectively.

The model size is a key metric that is limited by the onboard computers with restricted storage resources. As shown in Table 6, compared to the original model, the outcome of the designed SFM reduced the model size by 5.4 MB with almost no loss of accuracy.

Besides, the memory overheads of all the models are no more than 100 MB, which fits within the supported maximum synchronous dynamic random-access memory (SDRAM) of 512 MB for the current onboard computers [42]. This shows our model's superiority to state-of-the-art saliency detection models with several hundreds of megabytes [34,45] when deployed on SDRAM-limited devices.

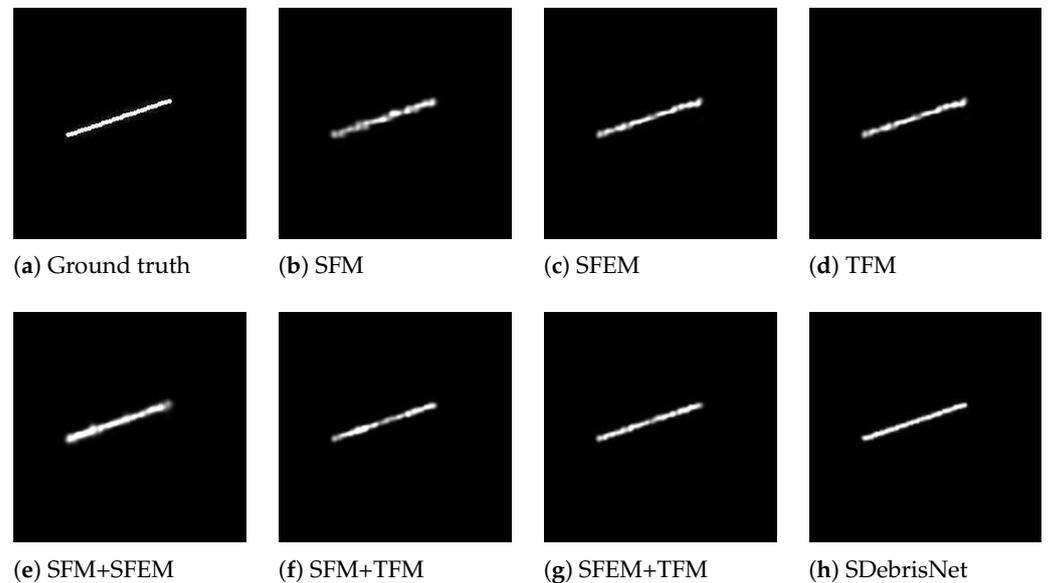


Figure 8. Saliency detection results of SDebrisNet without different components: all of the frames are collapsed by a max operator.

Table 6. The ablation study of SDebrisNet on the SDD dataset.

| SFM | SFEM | TFM | Model Size (MB) | $F_{\beta} \uparrow$ | $MAE \downarrow$ | $CE \downarrow$ |
|-----|------|-----|-----------------|----------------------|------------------|-----------------|
| | | | 48.44 | 0.724 | 0.143 | 1.316 |
| ✓ | | | 43.04 | 0.711 | 0.158 | 1.407 |
| | ✓ | | 74.32 | 0.729 | 0.138 | 0.946 |
| | | ✓ | 69.00 | 0.871 | 0.057 | 0.615 |
| ✓ | ✓ | | 68.92 | 0.763 | 0.065 | 1.082 |
| ✓ | | ✓ | 63.60 | 0.979 | 0.036 | 0.387 |
| | ✓ | ✓ | 94.88 | 0.989 | 0.035 | 0.383 |
| ✓ | ✓ | ✓ | 89.48 | 0.988 | 0.035 | 0.384 |

We compared the proposed method with the recent saliency detection models in terms of model size and computational efficiency. The results are listed in Table 7. We can see that the current saliency detection models mainly take the ResNet as the backbone to extract the spatial features. The model size of the proposed method is no more than 90 M and achieves a good inference speed among state-of-the-art saliency detection models.

Table 7. Comparison of the recent saliency detection methods.

| Methods | Backbone | Model Size (MB) | Inference Speed (FPS) | GPU |
|----------|-------------|-----------------|-----------------------|------------|
| MGA [59] | ResNet-101 | 350 | 14 | RTX 1080TI |
| DCF [60] | ResNet-101 | 274 | 28 | RTX 2080TI |
| STM [45] | ResNet-101 | 194 | 100 | RTX 2080TI |
| Ours | MobileNetV3 | 89 | 77 | RTX 2080TI |

5.4. Robustness Analysis

To assess the robustness of the proposed method, SDebrisNet was tested on the SDD dataset with different SNRs. The noise in a space surveillance platform mainly includes

thermal noise, shot noise, dark current noise, and stray light noise, which could affect the accuracy of space debris saliency detection. The first three noises are caused by the statistical nature of photodetection or photodiode in sensor systems [61], which can be represented as the Gaussian white noise model generally. The stray light noise results from accidental perturbations of diffuse reflection from the Earth, Moon, and other nebulae [62]. It also can be modeled using two Gaussian function dimensions. The detection results of SDebrisNet on the SDD dataset with different SNRs are shown in Table 8. Based on the test results, it can be observed that the centroid error increases as the SNR decreases in this experiment, and the false alarm rate also increases accordingly. The centroid error obtained by our method does not exceed 0.9 pixels in this experiment.

Table 8. Detection results of SDebrisNet on the SDD dataset with different SNRs.

| SNR | CE ↓ | P_d ↑ | P_f ↓ |
|-----|-------|---------|---------|
| 10 | 0.312 | 1.000 | 0.000 |
| 5 | 0.312 | 1.000 | 0.000 |
| 1 | 0.359 | 1.000 | 0.009 |
| 0.8 | 0.361 | 1.000 | 0.011 |
| 0.6 | 0.371 | 1.000 | 0.015 |
| 0.4 | 0.477 | 0.991 | 0.022 |
| 0.2 | 0.749 | 0.982 | 0.027 |
| 0.1 | 0.896 | 0.969 | 0.036 |

5.5. Test on Real Video Sequences

To further investigate the applicability of the proposed method, it was also tested on two real video sequences. The first video was acquired by a Newtonian telescope. It included two space debris named Iridium-33 and Cosmos-2251. The second video was obtained by the TJO telescope at the Observatorio Astronómico del Montsec (OAdM). Both videos were taken with the sidereal tracking mode. The tracks of the space objects are shown in Figure 9. The details of the observed space objects and optical sensors are shown in Table 9. In a practical circumstance, there is significantly more noise, intensity variations, and imaging artifacts in the real video, which increases the difficulty of the problem. Sub-frames were selected from the original frame size to show the above challenges, as shown in Figure 10. Since there are no known ground truth pixel coordinates of the space objects, only the qualitative test results are presented in this section. Part of the centroid coordinate detection results are shown in Figure 10. We can see that the intensity of Iridium-33 varies significantly in the continuous 3 frames (frame 340–frame 342). However, the space debris could still be detected by SDebrisNet. The second space debris Cosmos-33 is undetected in frame 435 due to the extremely low intensity caused by the tumbling characteristics. Frame 74 contains a striped white bright line called the smear-tailing phenomenon. It forms as the CCD is capturing a high-brightness point light source, which may affect the accuracy of detecting space debris in video images. Frame 320 has a lower SNR than frame 74, while frame 545 contains even more serious stray light noise compared to frame 320. The space object shows as a streak in frame 695. It is because the observation platform jitters in the horizon direction. Nevertheless, our method could work well in all of the difficulties in video 2.

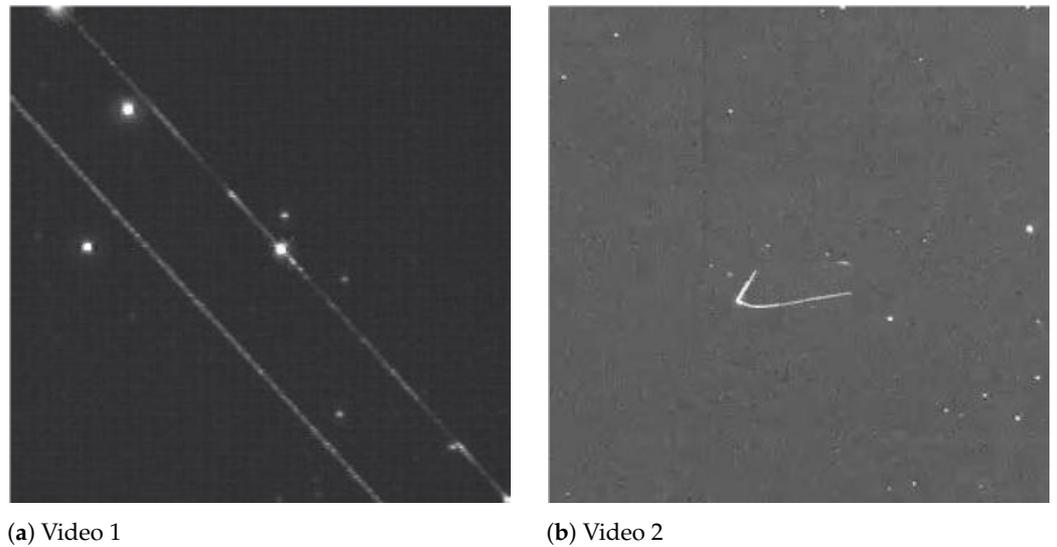


Figure 9. Two real video datasets: All of the frames are collapsed by a max operator. Video 1 includes two space debris with linear tracks. Video 2 includes one space object with a curved track.

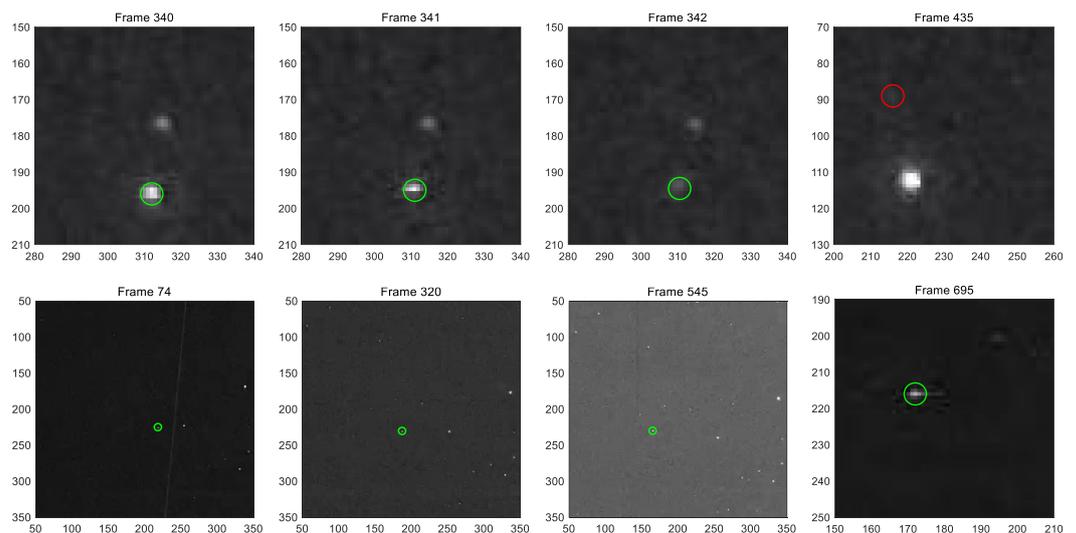


Figure 10. Example of the detected centroid on real video sequence 1 (top row) and video sequence 2 (bottom row). The detected centroid coordinates and missed detections are marked in green circles and red circles, respectively.

Table 9. Details of the real video sequences of the space objects.

| Video Number | Video 1 | Video 2 |
|--------------|-----------------------------------|--------------------------|
| Object name | Iridium-33 and Cosmos-2251 debris | Gaia spacecraft |
| Object size | 36–64 pixels | 6–64 pixels |
| Object speed | 1.48 pixels/frame | 0.16 pixels/frame |
| Tracks type | line | curve |
| Noise type | Gaussian | Gaussian and stray light |
| Frame rate | 30 frame/s | 25 frame/s |
| Frame number | 763 | 788 |
| Frame size | 720 pixels × 480 pixels | 360 pixels × 360 pixels |

6. Conclusions

In this paper, a novel spatial–temporal saliency-based approach was proposed for space debris detection. The proposed approach uses deep learning techniques to output

the saliency maps and compute the centroid coordinates of the space debris from satellite video sequences. The approach achieves end-to-end saliency detection of space debris without multiple steps of traditional methods. Moreover, as the model can generalize well to unseen space debris, due to the self-contained training samples, it does not require motion information from the space debris to be known a priori.

First, a lightweight spatial feature network was established to enable the inference model suitable for deployment on onboard devices with limited storage space. Based on the lightweight backbone, a spatial feature enhancement network was created by capturing both the image-level global context and the multi-scale spatial context. The experiment results show that the spatial feature enhancement network can extract more complex semantic features and complete shape features of small objects. Most importantly, a temporal feature extraction network was introduced by establishing the pair-wise relationships among feature elements in consecutive frames. This enables our method to detect space debris with a curved motion track due to the highly correlated temporal information existing in consecutive frames of satellite video. In addition, a public satellite video dataset for space debris detection was created to evaluate the proposed method. The motion speed and direction of space debris and the diameter of space debris and SNR are considered in the dataset. The results of the detection performance experiments show that both the detection probability and the false alarm rate of the space debris extracted by our method are much greater than those of both the single-frame-based method and the multi-frame-based method. In the sensitivity analysis experiment, the best video clip length used for each batch in the training stage is recommended as 4. The ablation study demonstrates that the model size of our method is largely reduced compared to the current saliency detection models. Moreover, we also tested the robustness of the proposed method on the SDD dataset with different SNRs. When the SNR of space debris is approximately 0.1, the centroid error of the space debris extracted by our method does not exceed 0.9 pixels. Finally, we tested our method on two real video datasets with more challenges. It showed that our method can work well in the space background with significant intensity variations, curved motion tracks, and stray light noise. The advantages of the proposed method can be summarized as follows. (1) The proposed method could achieve end-to-end space debris detection without multiple preprocessing steps. (2) It can detect both the linear and curved movements of space debris without a priori information. (3) It can detect space debris with a low SNR of 0.1. The main disadvantage of the proposed method is that it needs to create a video dataset containing different SNR and space debris movements.

The next step in our research will involve testing our method on an embedded platform with the same computational efficiency and storage capacity as the onboard computers.

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

Funding: This work was supported by the National Natural Science Foundation of China (grant number U2033201) and by funding from the Outstanding Doctoral Dissertation in the NUAU (grant NO. BCXJ19-11).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data (including the SDD datasets used to support the findings of this study) are available from the corresponding author upon request.

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

References

1. Li, Z.; Wang, Y.D.; Zheng, W. Space-based optical observations on space debris via multipoint of view. *Int. J. Aerosp. Eng.* **2020**, *2020*, 8328405. [[CrossRef](#)]
2. Zhang, J.R.; Shi, A.R.; Yang, K.Y. Dynamics of Tethered-Coulomb Formation for Debris Deorbiting in Geosynchronous Orbit. *J. Aerosp. Eng.* **2022**, *35*, 04022015. [[CrossRef](#)]
3. NASA Orbital Debris Program Office. Monthly object type charts by number and mass. *Orbital Debris Q. News* **2022**, *26*, 1–10.
4. Fang, Y.W.; Pan, J.; Luo, Y.J.; Li, C.W. Effects of deorbit evolution on space-based pulse laser irradiating centimeter-scale space debris in LEO. *Acta Astronaut.* **2019**, *165*, 184–190. [[CrossRef](#)]
5. Yang, W.; Zhao, Y.; Liu, M.; Liu, D. Method of space object detection by wide field of view telescope based on its following error. *Opt. Express* **2021**, *29*, 35348–35365. [[CrossRef](#)]
6. Diprima, F.; Santoni, F.; Piergentili, F.; Fortunato, V.; Abbattista, C.; Amoroso, L. Efficient and automatic image reduction framework for space debris detection based on GPU technology. *Acta Astronaut.* **2018**, *145*, 332–341. [[CrossRef](#)]
7. Fitzmaurice, J.; Bédard, D.; Lee, C.H.; Seitzer, P. Detection and correlation of geosynchronous objects in NASA's Wide-field Infrared Survey Explorer images. *Acta Astronaut.* **2021**, *183*, 176–198. [[CrossRef](#)]
8. Virtanen, J.; Poikonen, J.; Säntti, T.; Komulainen, T.; Torppa, J.; Granvik, M.; Muinonen, K.; Pentikäinen, H.; Martikainen, J.; Näränen, J.; et al. Streak detection and analysis pipeline for space-debris optical images. *Adv. Space Res.* **2016**, *57*, 1607–1623. [[CrossRef](#)]
9. Do, H.N.; Chin, T.J.; Moretti, N.; Jah, M.K.; Tetlow, M. Robust foreground segmentation and image registration for optical detection of GEO objects. *Adv. Space Res.* **2019**, *64*, 733–746. [[CrossRef](#)]
10. Jiang, P.; Liu, C.; Yang, W.; Kang, Z.; Fan, C.; Li, Z. Automatic extraction channel of space debris based on wide-field surveillance system. *Npj Microgravity* **2022**, *8*, 1–10. [[CrossRef](#)]
11. Brad, S.; Christopher, D.; Charles, P.; Jose, R.; Megan, J. Multi-stage astrometric image processing using stellar feedback. *Adv. Astronaut. Sci.* **2021**, *175*, 1–13.
12. Kouprianov, V. Distinguishing features of CCD astrometry of faint GEO objects. *Adv. Space Res.* **2008**, *41*, 1029–1038. [[CrossRef](#)]
13. Sun, Q.; Niu, Z.; Wang, W.; Li, H.; Luo, L.; Lin, X. An adaptive real-time detection algorithm for dim and small photoelectric GSO debris. *Sensors* **2019**, *19*, 4026. [[CrossRef](#)]
14. Sun, R.Y.; Zhan, J.W.; Zhao, C.Y.; Zhang, X.X. Algorithms and applications for detecting faint space debris in GEO. *Acta Astronaut.* **2015**, *110*, 9–17. [[CrossRef](#)]
15. Uetsuhara, M.; Hanada, T.; Yamaoka, H.; Fujiwara, T.; Yanagisawa, T.; Kurosaki, H.; Kitazawa, Y. Detection of faint GEO objects using population and motion prediction. In Proceedings of the 11th Annual Advanced Maui Optical and Space Surveillance Technologies Conference, Maui, HI, USA, 14–17 September 2010.
16. Montanaro, A.; Ebisuzaki, T.; Bertaina, M. Stack-CNN algorithm: A new approach for the detection of space objects. *J. Space Saf. Eng.* **2022**, *9*, 72–82. [[CrossRef](#)]
17. Yanagisawa, T.; Nakajima, A.; Kimura, T.; Isobe, T.; Futami, H.; Suzuki, M. Detection of small GEO debris by use of the stacking method. *Trans. Jpn. Soc. Aeronaut. Space Sci.* **2002**, *44*, 190–199. [[CrossRef](#)]
18. Yanagisawa, T.; Kurosaki, H.; Banno, H.; Kitazawa, Y.; Uetsuhara, M.; Hanada, T. Comparison between four detection algorithms for GEO objects. In Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference, Maui, HI, USA, 11–14 September 2012; Volume 1114, p. 9197.
19. Torteeka, P.; Gao, P.Q.; Shen, M.; Guo, X.Z.; Yang, D.T.; Yu, H.H.; Zhou, W.P.; Zhao, Y. Space debris tracking based on fuzzy running Gaussian average adaptive particle filter track-before-detect algorithm. *Res. Astron. Astrophys.* **2017**, *17*, 18. [[CrossRef](#)]
20. Uetsuhara, M.; Ikoma, N. Faint debris detection by particle based track-before-detect method. In Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference, Maui, HI, USA, 9–12 September 2014; p. E54.
21. Li, M.; Yan, C.; Hu, C.; Liu, C.; Xu, L. Space target detection in complicated situations for wide-field surveillance. *IEEE Access* **2019**, *7*, 123658–123670. [[CrossRef](#)]
22. Xi, J.; Wen, D.; Ersoy, O.K.; Yi, H.; Yao, D.; Song, Z.; Xi, S. Space debris detection in optical image sequences. *Appl. Opt.* **2016**, *55*, 7929–7940. [[CrossRef](#)]
23. Blostein, S.D.; Huang, T.S. Detecting small, moving objects in image sequences using sequential hypothesis testing. *IEEE Trans. Signal Process.* **1991**, *39*, 1611–1629. [[CrossRef](#)]
24. Rambaux, N.; Vaubaillon, J.; Lacassagne, L.; Galayko, D.; Guignan, G.; Birlan, M.; Boisse, P.; Capderou, M.; Colas, F.; Deleflie, F.; et al. Meteorix: A cubesat mission dedicated to the detection of meteors and space debris. In Proceedings of the 1st ESA NEO and Debris Detection Conference, Darmstadt, Germany, 22–24 January 2019; pp. 1–9.
25. Sun, T.; Xing, F.; Wang, X.; Li, J.; Wei, M.; You, Z. Effective star tracking method based on optical flow analysis for star trackers. *Appl. Opt.* **2016**, *55*, 10335–10340. [[CrossRef](#)] [[PubMed](#)]
26. Fujita, K.; Hanada, T.; Kitazawa, Y.; Kawabe, A. A debris image tracking using optical flow algorithm. *Adv. Space Res.* **2012**, *49*, 1007–1018. [[CrossRef](#)]
27. Tao, J.; Cao, Y.; Zhuang, L.; Zhang, Z.; Ding, M. Deep Convolutional Neural Network Based Small Space Debris Saliency Detection. In Proceedings of the 2019 25th International Conference on Automation and Computing (ICAC), Lancaster, UK, 5–7 September 2019; pp. 1–6. [[CrossRef](#)]

28. Zhang, X.; Xiang, J.; Zhang, Y. Space Object Detection in Video Satellite Images Using Motion Information. *Int. J. Aerosp. Eng.* **2017**, *2017*, 1024529. [[CrossRef](#)]
29. Yang, X.; Li, F.; Lu, M.; Xin, L.; Lu, X.; Zhang, N. Moving Object Detection Method of Video Satellite Based on Tracking Correction Detection. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *3*, 701–707. [[CrossRef](#)]
30. Steckling, M.; Renner, U.; Röser, H.P. DLR-TUBSAT, qualification of high precision attitude control in orbit. *Acta Astronaut.* **1996**, *39*, 951–960. [[CrossRef](#)]
31. Feng, J.; Zeng, D.; Jia, X.; Zhang, X.; Li, J.; Liang, Y.; Jiao, L. Cross-frame keypoint-based and spatial motion information-guided networks for moving vehicle detection and tracking in satellite videos. *ISPRS J. Photogram. Remote Sens.* **2021**, *177*, 116–130. [[CrossRef](#)]
32. Xiao, A.; Wang, Z.; Wang, L.; Ren, Y. Super-resolution for “Jilin-1” satellite video imagery via a convolutional network. *Sensors* **2018**, *18*, 1194. [[CrossRef](#)]
33. Jabir, B.; Falih, N.; Rahmani, K. Accuracy and Efficiency Comparison of Object Detection Open-Source Models. *Int. J. Online Biomed. Eng.* **2021**, *17*. [[CrossRef](#)]
34. Chen, Y.W.; Jin, X.; Shen, X.; Yang, M.H. Video Salient Object Detection via Contrastive Features and Attention Modules. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 1320–1329.
35. Xu, B.; Liang, H.; Ni, W.; Gong, W.; Liang, R.; Chen, P. Learning Video Salient Object Detection Progressively from Unlabeled Videos. *arXiv* **2022**, arXiv:2204.02008.
36. Zhao, W.; Zhang, J.; Li, L.; Barnes, N.; Liu, N.; Han, J. Weakly supervised video salient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 16826–16835.
37. Su, Y.; Deng, J.; Sun, R.; Lin, G.; Wu, Q. A Unified Transformer Framework for Group-based Segmentation: Co-Segmentation, Co-Saliency Detection and Video Salient Object Detection. *arXiv* **2022**, arXiv:2203.04708.
38. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 8–22 June 2018; pp. 7794–7803.
39. Yan, P.; Li, G.; Xie, Y.; Li, Z.; Wang, C.; Chen, T.; Lin, L. Semi-supervised video salient object detection using pseudo-labels. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7284–7293.
40. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
41. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
42. Blacker, P.; Bridges, C.P.; Hadfield, S. Rapid prototyping of deep learning models on radiation hardened cpus. In Proceedings of the 2019 NASA/ESA Conference on Adaptive Hardware and Systems (AHS), Colchester, UK, 22–24 July 2019; pp. 25–32.
43. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–November 2019; pp. 1314–1324.
44. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
45. Zhao, X.; Liang, H.; Li, P.; Sun, G.; Zhao, D.; Liang, R.; He, X. Motion-aware Memory Network for Fast Video Salient Object Detection. *arXiv* **2022**, arXiv:2208.00946.
46. Hu, P.; Ramanan, D. Finding tiny faces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 951–959.
47. Hong, M.; Li, S.; Yang, Y.; Zhu, F.; Zhao, Q.; Lu, L. SSPNet: Scale selection pyramid network for tiny person detection from UAV images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [[CrossRef](#)]
48. Mahadevan, S.; Athar, A.; Ošep, A.; Hennen, S.; Leal-Taixé, L.; Leibe, B. Making a case for 3d convolutions for object segmentation in videos. *arXiv* **2020**, arXiv:2008.11516.
49. Wang, H.; Mu, N.; Zhang, Y. Video Salient Object Detection Network with Bidirectional Memory and Spatiotemporal Constraints. In Proceedings of the 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, Australia, 17–20 October 2021; pp. 2781–2786. [[CrossRef](#)]
50. Liu, J.; Wang, J.; Wang, W.; Su, Y. DS-Net: Dynamic spatiotemporal network for video salient object detection. *Digital Signal Process.* **2022**, *130*, 103700. [[CrossRef](#)]
51. Gu, Y.; Wang, L.; Wang, Z.; Liu, Y.; Cheng, M.M.; Lu, S.P. Pyramid constrained self-attention network for fast video salient object detection. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 10869–10876.
52. Ji, G.P.; Chou, Y.C.; Fan, D.P.; Chen, G.; Fu, H.; Jha, D.; Shao, L. Progressively normalized self-attention network for video polyp segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France, 27 September–1 October 2021; pp. 142–152.
53. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [[CrossRef](#)]

54. Achanta, R.; Hemami, S.; Estrada, F.; Susstrunk, S. Frequency-tuned salient region detection. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1597–1604.
55. Hou, Q.; Cheng, M.M.; Hu, X.; Borji, A.; Tu, Z.; Torr, P.H. Deeply supervised salient object detection with short connections. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3203–3212.
56. Fan, D.P.; Cheng, M.M.; Liu, Y.; Li, T.; Borji, A. Structure-measure: A new way to evaluate foreground maps. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4548–4557.
57. Guo, X.; Chen, T.; Liu, J.; Liu, Y.; An, Q. Dim Space Target Detection via Convolutional Neural Network in Single Optical Image. *IEEE Access* **2022**, *10*, 52306–52318. [[CrossRef](#)]
58. Liu, D.; Chen, B.; Chin, T.J.; Rutten, M.G. Topological sweep for multi-target detection of geostationary space objects. *IEEE Trans. Signal Process.* **2020**, *68*, 5166–5177. [[CrossRef](#)]
59. Li, H.; Chen, G.; Li, G.; Yu, Y. Motion guided attention for video salient object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 7274–7283.
60. Zhang, M.; Liu, J.; Wang, Y.; Piao, Y.; Yao, S.; Ji, W.; Li, J.; Lu, H.; Luo, Z. Dynamic context-sensitive filtering network for video salient object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1553–1563.
61. Hui, R. *Introduction to Fiber-Optic Communications*; Academic Press: London, UK, 2019; pp. 125–154.
62. Park, J.O.; Jang, W.K.; Kim, S.H.; Jang, H.S.; Lee, S.H. Stray light analysis of high resolution camera for a low-earth-orbit satellite. *J. Opt. Soc. Korea* **2011**, *15*, 52–55. [[CrossRef](#)]

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