On-orbit Moving Target Detection Method Based on Dual Images from Taijing-IV 01 Satellite

Wei Li[1], Cankun Yang[1,2*], Xiaojuan Li[1,2], Lianwang Yin[1,2], Meizhi Cheng[1]

[1] College of Resource, Environment, and Tourism, Capital Normal University, Beijing, China - 2220902132@cnu.edu.cn, yangck@cnu.edu.cn, lixiaojuan@cnu.edu.cn, YinLW2011@gmail.com, 2190902202@cnu.edu.cn
[2] Key Laboratory of 3D Information Acquisition and Application, Capital Normal University, Beijing, China - yangck@cnu.edu.cn, lixiaojuan@cnu.edu.cn, YinLW2011@gmail.com

Keywords: Motion Target Detection, Frame Difference, YOLO, On-orbit Processing.

Abstract

With the evolution of satellite video technology, the domain has garnered increasing attention. Concurrently, advancements in deep learning have yielded numerous outcomes in target detection. This paper introduces a novel method for detecting moving targets, offering a broader detection range compared to traditional satellite video techniques, facilitating orbital target recognition from dual panchromatic image strips. Our experimental setup on the Taijing-IV 01 satellite, launched on February 27, 2022, successfully acquired two image strips separated by one second. These strips contain speed and directional information of moving objects, extractable through the frame differencing technique. We propose combining frame differencing with lightweight deep learning for target detection, extracting regions of interest (ROIs) to focus on areas with potential moving targets. This approach reduces the workload of whole-image target detection, decreasing data processing volume by 89%. By optimizing the YOLOv8 network and using techniques like feature map fusion of low-level and high-resolution features, we enhance sensitivity to small targets. Consequently, the model size is reduced by 79%, the mean Average Precision (mAP) increases by approximately 1.8% and 4.5%, and detection speed rises by 26%. This method introduces a new paradigm in remote sensing data services, facilitating rapid acquisition and real-time transmission of positions and image information of moving targets to the ground. This significantly reduces bandwidth requirements for transmitting remote sensing information, presenting a novel strategy for data acquisition and processing in large-scale Earth observation systems and geoscientific applications.

1. Introduction

Moving target detection is one of computer vision's most popular research directions. It is usually used as a pre-processing step of advanced visual tasks such as target detection, target tracking, and pose estimation to identify the moving target region of interest or focus in the video. It is widely used in the military, smart city, intelligent transportation, Self-driving, and many other fields, making moving target detection more attractive. In recent years, the rapid development of space technology and the optimization of sensor performance have driven the vigorous development of space remote sensing. With the development of low-Earth orbit satellites, video satellites with the ability to observe continuously have attracted more and more attention from experts and scholars. Moving target information extraction for satellite images has become a hot topic in scholars' research. High-resolution satellite images cover a wide range of areas and contain rich information, which can effectively monitor traffic conditions of multiple sections simultaneously, saving labour and cost. The extensive coverage of satellite remote sensing imagery provides the ability to analyse many vehicle movements, which is especially suitable for areas lacking cameras. Further, satellites for real-time moving target detection in orbit will provide more efficient and automated services that will significantly facilitate the development of applications in areas such as military reconnaissance and intelligent transportation.

Optical and synthetic aperture radar (SAR) data are two common types of remote sensing data. Regarding target detection, SAR/GMTI can observe moving targets in any weather and throughout the day. Increasing the length of the radar antenna or the number of satellites can achieve a more extended observation baseline, making it more suitable for slow-moving target detection. However, the above approach raises the issue of increased manufacturing cost and difficulty. Optical remote sensing images can use the rich spectral texture information in radiation features to achieve identification, classification, and interpretation. Many scholars have used satellite high-resolution multispectral and SAR images to detect, track and estimate the velocity of moving targets. These studies cleverly exploited the time delay between imaging sensors (typically less than 0.3 s) to detect the difference between two images and achieve moving target detection. However, this approach can only extract fast targets, and the detection and matching methods are mostly semi-automatic and require highly subjective manual parameter settings. Obtaining instantaneous dynamic information of targets is challenging for single-channel panchromatic images with relatively high resolution but lacking temporal differences. It can only be used to study target recognition algorithms. Video data from satellites such as SkySat and Jilin-1 have solved the problem of continuous observation of moving targets. However, on the one hand, video capture requires the satellite to perform gaze maneuvers, which limits the imaging range; on the other hand, redundant frames of video data will increase storage consumption and make real-time processing and transmission more difficult.

Although a series of previously mentioned moving target detection studies have achieved some superior results in slow-moving target detection, the following challenges still exist.

(1) Multispectral or hyperspectral satellite sensors are designed to image synchronously to ensure fusion accuracy. However, a short time difference for images means that detecting slow-moving targets is complicated.
(2) In contrast to panchromatic images, spectral images often provide lower resolution, which makes small target detection more challenging.
(3) Satellite video covers a fixed area with a large amount of data, which puts pressure on data storage and internal transmission.
(4) On-orbit processing payloads process remote sensing data in nearly real-time, creating high demands on processing efficiency. Directly transplanting the target detection algorithm of industrial security camera video to the on-orbit hardware platform is unsuitable because the power consumption and performance of the on-orbit hardware platform are limited. Inappropriate algorithms would consume too many hardware resources.

We notice that target detection algorithms are widely used in industrial security videos. At the same time, a large number of researchers have carried out a lot of improvement work for both efficiency and accuracy. Comparing industrial and on-orbit environments, we find that industrial target detection algorithms still need to be developed for the embedded environment and model lightweight and efficiency enhancement. These problems inspired us to combine the dual-linear image frame difference method and target detection algorithm with improving the efficiency and accuracy of target detection and making it more suitable for the on-orbit processing environment limitations.

The dual-strip imaging system principle employed in this study involves the CMOS sensor within the satellite camera capturing dual-strip images through two open window regions. With thousands of lines separating these windowed areas on the CMOS, the resulting imaging time delay between the two strips averages approximately 1 second. As a result, the two strips contain diverse spatial information about the same feature, facilitating the generation of an initial saliency map essential for moving target detection. Subsequently, during the satellite's push-broom working mode, the entire strip undergoes target detection and classification using a YOLO model accelerated by FPGA and GPU. Throughout this process, image slices derived from the saliency map significantly reduce the data input required for the YOLO model. In the context of on-orbit processing tasks, the hardware and software requirements are predominantly focused on efficiency and resource utilization due to the significant cost escalation associated with increased hardware weight, size, and power consumption. Consequently, considering the frame difference method for completing data processing tasks is a more practical option for dual-strip images. However, it is imperative to note that the selection of thresholds during processing significantly impacts the recall and false alarm rate of moving target detection. Moreover, in comparison to satellite remote sensing, the reduction in target size to a few tens of pixels results in decreased detection accuracy. Additionally, the hundred-fold increase in data volume of remote sensing images compared to video frame images poses a considerable challenge to hardware design. This paper proposes an on-orbit moving object detection method for dual-strip imagery, integrating the frame difference method and deep learning, to achieve the detection and classification of moving targets with hardware acceleration design. Leveraging saliency maps generated from dual-strip images significantly reduces the data input to the model while enhancing efficiency through preprocessing and first-level data production by filtering valuable regions.

The results of this paper have been successfully validated in orbit using a payload named Frog Eyes. The Taijing IV 01 satellite was launched on February 27, 2022, carrying the Frog Eyes payload. Following the successful launch of this mission, the payload captured panchromatic dual-strip images to achieve full-range moving object detection. On-orbit processing was utilized instead of the traditional process, which typically involves mission planning, satellite imaging, satellite-to-ground data transfer, and ground processing. The time taken from raw data to target information, as well as the volume of resulting data transfer, determine whether remote sensing image target detection can meet the requirements of applications, such as military and intelligent transportation.

2. Related Work

2.1 Moving Target Detection Based on Multi-frame Image

The research of moving target detection in satellite remote sensing images is mainly divided into moving target recognition based on single-frame images and moving target detection based on multi-frame sequence images. Most research on moving target detection focus on vehicle detection, while the researches on ships and aircraft are relatively few. When researchers started to propose using satellite remote sensing imagery for moving target detection, they mainly chose QuickBird, WorldView, and high-resolution SAR radar data for their work. Research on moving target detection in single-frame images has been abundant, and regular algorithms include threshold segmentation, template matching, and object-oriented and artificial neural networks. Motion target detection methods for multi-frame data are inherited from traditional video target detection algorithms, with several improved versions. Commonly used algorithms include the background difference method, optical flow method, and inter-frame difference method. The inter-frame difference algorithm has low complexity, small computational effort, and is insensitive to image exposure changes so that it can be applied to target detection in most scenes. The inter-frame difference is a moving target detection method based on the difference operation of sequential images of two or three consecutive frames. Researchers proposed a symmetric differential moving target detection algorithm based on local clustering segmentation (Cao et al., 2015). This algorithm, when applied to infrared aerial sequence images, can detect relatively complete "slow" moving targets with small motion changes between adjacent frames.

A proposal was made to employ the three-frame difference method for detecting moving targets with morphological constraint processing, yielding commendable detection results (Yu, 2016). The algorithm was subsequently embedded into DSP for real-time processing. Serving as an enhanced version of the two-frame differencing method, the three-frame differencing method distinguishes the front and back frames from the reference frame. It sets a threshold to filter regions containing significant values from the result data, thereby reducing the false alarm rate of detection results. Subsequently, the background modeling method was proposed to obtain the deviation area of the moving target, denoise the target based on actual area constraints, improve the detection rate, and enable satellite video moving target detection. The algorithm demonstrated a higher detection rate compared to the earlier three-frame difference method. In 2018, a target detection algorithm was proposed that fuses background differencing with inter-frame differencing, combining the common region of motion target changes with the results extracted by the background difference method. This method removes background edge noise and single-point noise to enhance detection accuracy and quality (Yuan et al., 2018). A deep frame difference convolutional neural network (DFDCNN) was designed. The network comprises DifferenceNet and AppearanceNet, capable of detecting moving objects in complex scenes. Preservation of multi-scale spatial information through multi-scale feature map fusion and progressive upsampling improves the sensitivity of DFDCNN networks to small targets (Ou et al., 2020). In 2021, a tracking algorithm of scale adaptive kernel correlation filter combined with a frame difference method was proposed to address the unsatisfactory tracking effect of kernel correlation filter in complex scenes (Liu et al., 2021). An improved intelligent video tracking and detection algorithm was also proposed to enhance ordinary video moving target tracking technology in real time (Liu et al., 2021). To reduce computational effort, the frame difference method is used to capture the region of the moving target. Then, the optimal estimation points matching technique and the uniform flow technique are utilized to construct the optical flow field for target detection and tracking.

2.2 Satellite Image Target Detection Based on Deep Learning

Deep learning techniques are now achieving great success in computer vision. A convolutional neural network can automatically learn features, extract rich target features for subsequent detection, and avoid the difficulty of feature selection with good detection performance (Qi, 2020). The YOLO series is the most popular target detection algorithm, but there are still many problems transferring it to remote sensing image target detection. A large amount of remote sensing image data and the broad coverage area result in smaller aircraft, ships, and vehicles occupying fewer pixels on the image. A single tiny target contains little pixel information. These small targets extract minimal practical information in target detection, and the detection recall rate is greatly affected. As for small target feature extraction by mesh convolutional neural networks, many models usually go through several sampling operations to increase the perceptual field and continuously reduce the dimensions to shrink the feature map and make the semantic information less. Sometimes the tiny targets are severely lost after the dimensionality reduction operation and cannot even be effectively transmitted to the target detector (Liang et al., 2021).

2.3 On-orbit Processing of Remote Sensing Satellite Image

With the continuous development of satellite technology, the types of remote sensing data have become more diverse, and users' needs have become more complex. The traditional satellite data acquisition and processing mode can no longer meet the high-efficiency information requirement of users. With the advent of continuous observation data such as satellite video, the data collected by the satellite increases several times per second, which is a massive challenge for satellite data storage and transmission. Based on this, it is essential to study onboard processing algorithms, which extract and transmit specific information and regions of interest from the original images and no longer transmit the raw data to the ground. Currently, China has several remote sensing satellites with on-orbit processing capability in space, including "Space Experiment 1", launched in September 2015, and "Tianzhi 1", launched in November 2018. Moreover, the launch of the "Luojia III" satellite with on-orbit processing capability is also on the agenda (Data and Application Centre for High Resolution Earth Observation System in Hubei, 2019). The Beijing Institute of Technology proposed the on-orbit real-time processing technology of space remote sensing for SAR satellite processing, strip mode imaging processing, and target detection and positioning (Liu, 2016). The Computer Network Information Centre of the Chinese Academy of Sciences designed a distributed and networked on-orbit processing system of space-based information (Yue et al., 2020).

3. Proposed Method

This paper presents a method to detect moving targets using two panchromatic strip images from a particular camera. The remote sensing camera has a built-in multimodal super COMS designed. The camera can acquire strip images and read and take pictures with flexible on-chip windowing. In Figure 1, we design two areas on the chip thousands of rows apart to take pictures of the Earth's surface and work in push-broom mode to acquire two panchromatic strip images with one-second intervals. The subsatellite ground resolution and the number of rows between open window regions determine this time difference. Thus, two images cover the same area but are imaged at different moments to acquire dynamic information from each time. Considering the characteristics of MS-COMS imaging, the panchromatic dual-strip satellite image moving target detection solution proposed in this paper accommodates both on-satellite and ground processing. The on-orbit processing module acquires the information and image slices of the moving target, which can transmit only a tiny amount of information during the satellite-ground transmission, then reproduce the features of the striped image by reconstruction algorithm on the ground. The on-orbit processing module mainly includes the following functions.

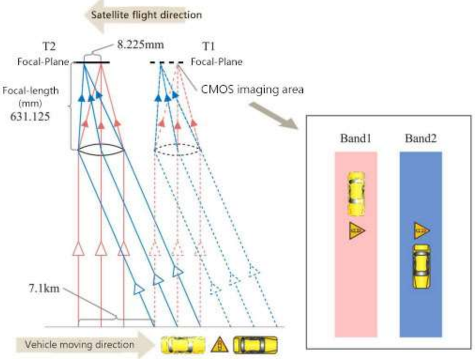

Figure 1. Schematic diagram of acquiring dual-band images of moving targets

(1) Extraction of the suspected target area;
(2) Classify and identify the suspected target;
(3) Moving target matching and motion parameter extraction.
This paper will focus on the part of in-orbit processing. Figure 2 shows the Technical Flow.

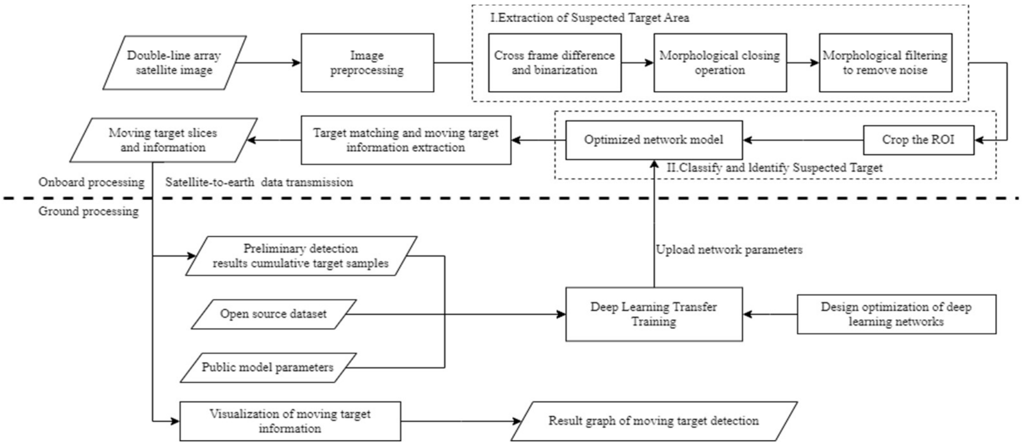

Figure 2. The flow of on-orbit moving target detection.

3.1 Extraction of Suspected Target Area

Since the signal response, image element damage, and dark current of the two CMOS regions are not precisely the same. It is necessary to perform a series of image pre-processing process to ensure the confidence of the motion target detection results. The image pre-processing process generally includes relative radiation correction, systematic geometric correction, and image registration to eliminate radiation and geometric errors. After pre-processing, the frame difference method can be selected to extract suspicious target regions from a pair of strip images. The frame difference method is the most commonly used method to detect inter-frame variations. It can compare the difference between two frames by pixel value in grayscale and then determine whether there is a moving target. The moving object's position at both moments can be obtained by cross-subtracting the images of the two moments. Equation 1 shows the theoretical formula of the frame difference method used in this paper.

$$\begin{cases} D_{1(x,y)} = f(x,y) - f_{k+1}(x,y) \\ D_{2(x,y)} = f_{k+1}(x,y) - f_k(x,y) \end{cases} \tag{1}$$

In the above equation, (x, y) is pixel position in the image, $D_{1(x,y)}$ is the pixel value of first frame image (T1) after calculation, $D_{2(x,y)}$ is the pixel value of second frame image (T2) after calculation, $f_k(x, y)$ is the pixel value of T1, $f_{k+1}(x, y)$ is the pixel value of T2.

After getting the frame difference results, choose a suitable threshold to binarize the frame difference results and further obtain the significance map. The foreground region in the saliency map indicates the significant grayscale difference between images during frame difference calculation, which can be regarded as the moving target region. The other regions are filled as a background. During processing, a threshold value needs to be determined to ensure a high detection rate, and this value can be set by experience and updated on-orbit by a command. Targets in the saliency map can be enhanced by a morphological closure operation, where target pixel expansion is followed by erosion. Figure 3 shows the schematic diagram of the morphological closure operation.

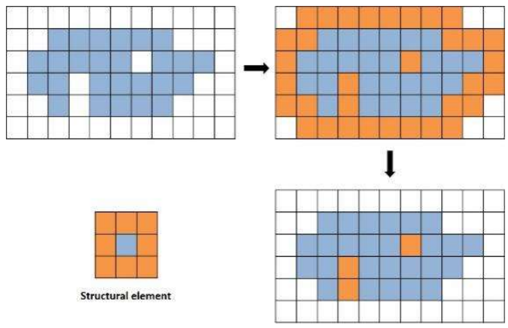

Figure 3. The schematic diagram of closed operations in morphology.

Subject to slow target speed or low image alignment accuracy and brightness difference images, boundary overlap between targets needs attention. After completing the inter-frame difference calculation, a threshold needs to be determined to binarize the image and project it into the X-Y plane, a value based on a priori knowledge of the minimum possible brightness of the target. Small holes or cracks appear within a single target in the saliency map. A morphological region growth operation fills these small holes and cracks, and then an erosion operation is performed on the swollen image to restore the original shape. The process introduces morphological information to achieve noise filterings, such as aspect ratio and the average of the gray value of the foreground. Linear noise, such as ground object boundaries, and single point noise, such as streaks, are removed.

3.2 Classify and Identify Suspected Target

To ensure the recall rate of target detection, Choosing a threshold value as small as possible during the frame difference method to obtain the saliency map is very important. As the two ends of the scale, this also has the disadvantage of generating a higher false alarm rate. We use a deep learning target detection algorithm to perform secondary recognition of the region confirmed by the saliency map and detect moving targets from the region, which can significantly reduce the false alarm rate. At the same time, it effectively solves the problem of incomplete target outline in saliency maps, and the target recognition on the original map can guarantee the integrity of the target outline. In order to reduce the data input to the model, it makes conditional judgments on the slice data. It selects only those slices containing the target location in the frame difference saliency map, which is also recorded as POI information for subsequent processing. Each slice is 512 pixels long and wide, and the model only reads the slice containing the target, which can significantly reduce the amount of data processed by the model and improve real-time performance.

YOLO(You Only Look Once) have developed rapidly, merging target determination and target identification and regressing directly in the output layer to generate detection frames and belonging classes, with excellent performance in terms of accuracy and efficiency. The YOLO network inference computation includes many convolutional operations, which are time-consuming and have a strong optimization potential. To address the issues of incomplete targets and increased false alarms during the moving target detection process, a lightweight motion target classification model based on deep learning is proposed to authenticate the results of the first stage. The algorithm is inspired by the YOLOv8s target detection model, with optimized capabilities for detecting small targets. Within the network architecture, a specialized prediction head for detecting tiny objects is added, which combines low-level, high-resolution feature maps to enhance sensitivity to small targets. Although this method increases computation and storage costs, it mitigates the negative impact of target scale variations. In terms of efficiency, model volume optimization and inference efficiency improvement are achieved through sparse training and pruning methods.

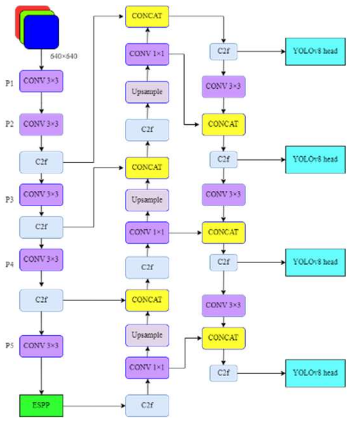

Figure 4. Network structure with improved optimization for small target detection

3.3 Estimate Motion Parameters of Target

The saliency map indicates the target's position at two moments, and by matching the features of the target at different moments, the paired target can estimate its motion parameters. It also needs to build a database of motion information of suspicious targets in two images separately, determine which two objects are the same based on feature information, and match them.

According to the target features in the label of the suspected target area obtained by the deep learning target detection, the following determination indicators are formulated:
(1) Type: Target type from target recognition box type.
(2) Target distance: The distance between two target positions. The centre of the outer rectangle is used to calculate the distance, and the upper limit is set according to the physical limit speed of different types of targets. It is assumed that the target moves with a constant velocity in 1s.
(3) Diff-area: in the satellite image, the target boundary is fuzzy, and the length and width of the moving target have about 1 pixel error.

$$(\mathrm{w}_1-1)\times(\mathrm{h}_1-1) \leq \mathrm{Area}_2 \leq (\mathrm{w}_1+\frac{\mathrm{w}_1}{10})\times(\mathrm{h}_1+\frac{\mathrm{h}_1}{10}) \tag{2}$$

In the formula, h_1 and w_1 are the length and width of the object at T1, $Area_2$ is the area of object at T2.

(4) The similarity of targets: The target detection yields two sets of individual target masks and associated target parameter information, including centroid position, area, mean grayscale, and standard deviation. To match the moving targets at two moments, the grayscale information of the moving targets at the first moment is utilized as much as possible to correctly match the moving targets in the second moment, forming a list of matched targets. The extracted information of the two sets of moving targets is regarded as two-dimensional lists, and the Euclidean distance is introduced to measure the similarity of parameter vectors between the two lists. The rationality of this method has been extensively discussed in many studies. A smaller Euclidean distance indicates a greater similarity in grayscale distribution between two moving targets, while a larger distance indicates greater differences between them.

$$\begin{cases} A = (A_{area}, A_{avg}, A_{std}) \\ B = (B_{area}, B_{avg}, B_{std}) \\ D_{Eu}(A,\ B) = \sqrt{(A_{area}-B_{area})^2 + (A_{avg}-B_{avg})^2 + (A_{std}-B_{std})^2} \end{cases} \tag{3}$$

Equation (3) represents the similarity between the feature vectors of the moving targets at time t and t+1. A denotes the feature vector of the target at time t, with attributes A_{area}, A_{avg}, and A_{std} representing the area, average grayscale value, and standard deviation of the pixel region, respectively. B denotes the feature vector of the target at time t+1, with attributes B_{area}, B_{avg} and B_{std} representing the area, average grayscale value, and standard deviation of the pixel region, respectively, $D_{Eu}(A,B)$ represents the similarity between feature vectors A and B described by the Euclidean distance.

The targets at the two moments are calculated according to the above indicators, and the successfully matched targets are entered into the matching information base. The algorithm deletes the object from the database when the best match occurs to prevent one-to-many matches in the results. After successfully matching the target, it is necessary to calculate its displacement, velocity and direction. The target displacement distance determines the velocity. The displacement can be estimated based on the position of the same target at two moments $(\mathrm{X}_1, \mathrm{Y}_1)$ and $(\mathrm{X}_2, \mathrm{Y}_2)$ and the image resolution R. Equation (4) below shows the process. The direction of motion θ is expressed as the angle between the displacement vector and the north direction. Judging the direction of motion requires calculating the position $(\mathrm{x}_1, \mathrm{y}_1)$ and $(\mathrm{x}_2, \mathrm{y}_2)$ of the same target at two moments. Formula (5) describes the calculation method.

$$D = R \times \sqrt{(X_1-X_2)^2 + (Y_1-Y_2)^2} \tag{4}$$

$$\theta = \begin{cases} \arctan\frac{|x_2-x_1|}{|y_2-y_1|} & x_2 > x_1, y_2 > y_1 \\ \pi - \arctan\frac{|x_2-x_1|}{|y_2-y_1|} & x_2 > x_1, y_2 < y_1 \\ \pi + \arctan\frac{|x_2-x_1|}{|y_2-y_1|} & x_2 < x_1, y_2 < y_1 \\ 2\pi - \arctan\frac{|x_2-x_1|}{|y_2-y_1|} & x_2 < x_1, y_2 > y_1 \end{cases} \tag{5}$$

4. Experiments

4.1 Classify and Identify Suspected Target

To validate the effectiveness of our method, we conducted experiments using dual-strip images captured by the Taijing-IV 01 satellite in orbit, with a spatial resolution of 2 meters per pixel. The imaging area was near the Port of Los Angeles, USA, and the imaging time was August 2022. The experimental setup utilized the NVIDIA AGX Xavier development board computing platform, which includes an NVIDIA Volta architecture GPU, an 8-core ARM64 CPU, 512 CUDA cores, and 64 Tensor cores. The experiment area is conducted over the area of the Port of Los Angeles in the United States, with the image dimensions of 5000×5000 pixels. The images contain numerous small and dim targets. Figure 5 shows a binary image with highlighted pixels as primary suspected targets. When cropping the ROI in the original image, there exist two methods, target-centred cropping, and grid cropping. The former method requires traversing the distances between targets and classifying them into discrete and aggregated types based on the distance between the geometric centres of each target. At each iteration, the current benchmark targets will be classified, and targets with a Euclidean distance of fewer than 64 pixels from the benchmark will be considered a cluster. In contrast, isolated benchmark targets without neighbours are considered discrete. In this way, there are slices containing a single target and those containing multiple targets. The grid cropping algorithm divides the map frame into square grids of 512 pixels, determines whether there are suspected targets in each grid separately, and crop out the grid areas containing suspected targets. As a comparison, target-centred cropping retains fewer grids and has a higher compression ratio but is high computationally. On the other side, grid cropping produces redundant grids to avoid missing targets on the grid divider, thus generating a more significant number of grids with lower compression ratios, but the computerization is small. We chose the grid cropping method with higher efficiency to better accommodate the hardware performance of in-orbit processing.

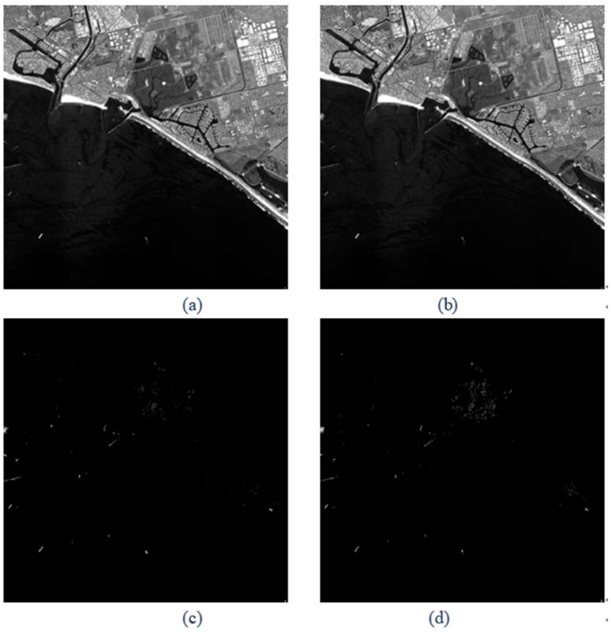

Figure 5. The result of moving target detection by frame difference method. (a) and (b) are images for testing, (c) and (d) are saliency map for frame difference and denoising results.

The applicability of this method was subsequently tested on images from the Taijing-IV 01 satellite with dimensions of 5000×5000 pixels, as shown in Figure 6. This is crucial for accurately interpreting the results. Table 1 presents quantitative statistics, including the total number of suspicious targets, the number of grid slices, the number of ROI slices, and the percentage of data reduction. Compared to grid cropping and target-centred cropping, the data volume of the two frames after slicing was effectively reduced, with reductions of 36%, 32%, 10%, and 7%, respectively. After cropping, the data input to the target recognition model decreased, improving the real-time performance of the algorithm.

	grids	suspected targets	ROI	ROI /grids	ROI/ suspected targets
Frame1	100	353	36	0.36	0.102
Frame2	100	422	32	0.32	0.071

Table 1. Experimental results of cropping.

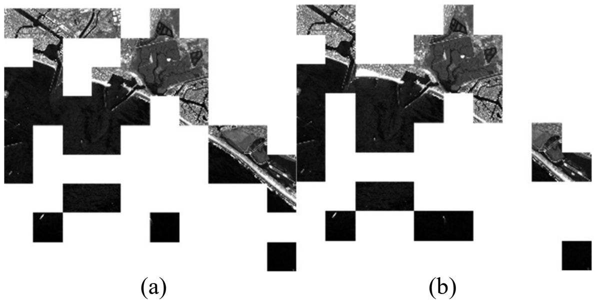

Figure 6. According to the location of the grid, the results of stitching the cropped regions of interest (ROIs) are shown. (a) and (b) represent slices of the moving target ROIs in the first and second frames, respectively.

4.2 Moving Target Detection Based on YOLO

We optimized the YOLOv5s and YOLOv8s networks for small object prediction heads and sparse pruning training, and compared the networks before and after optimization. The results, as shown in Figure 6, on the dataset collected by the Taijing-IV 01 satellite, indicate that our models YOLOv5s+mini+prune and YOLOv8s+mini+prune achieved desirable results across various metrics. In Table 2, for YOLOv5s, the mAP accuracy improved by 1.6% and 5.5% respectively compared to the original network before optimization; FPS increased by 56%; the model size decreased by 45.4%, and the parameter count reduced by 70.2%. As for YOLOv8s, the mAP accuracy improved by 1.8% and 4.5% respectively compared to the original network before optimization; FPS increased by 26%; the model size decreased by 79%, and the parameter count reduced by 80.1%. This enhancement in sensitivity to detecting small objects also led to an improvement in target detection efficiency.

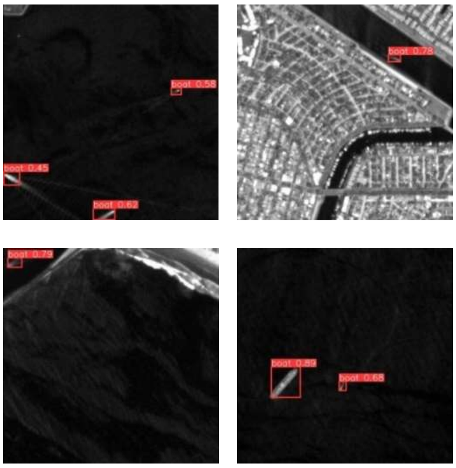

Figure 7. Object detection results based on YOLO

4.3 Results of Estimate Motion Parameters

In the motion target detection algorithm for dual-strip satellite images presented in this paper, the original image data was first cropped to extract suspected target regions before deep learning-based target detection. When performing the target algorithm based on object feature matching, mapping all target information back to the original image based on the cropping results before matching may lead to a decrease in algorithm efficiency. Since the time difference between the two images is only 1 second and the distance of target motion is minimal, the majority of targets remain in the same grid position before and after motion. To address this characteristic, improvements were made to the target tracking algorithm based on object feature matching. It prioritizes block-level target matching, followed by matching failed targets across the entire image, as illustrated in Figure 8 to calculate motion parameter results.

Model	mAP@0.5	mAP@0.5:0.95	FPS	Model Size(M)	Parameter
YOLOv5s	0.926	0.493	25	14.1	7046599
YOLOv5s+mini	0.931	0.457	18	15.5	7192244
YOLOv5s+mini+prune	0.941	0.520	39	7.7	2101396
YOLOv8s	0.935	0.598	35	22.5	11166544
YOLOv8s+mini	0.944	0.622	29	21.7	10627228
YOLOv8s+mini+prune	0.952	0.625	44	4.7	2218145

Table 2. Performance comparison small target prediction head and sparse pruning

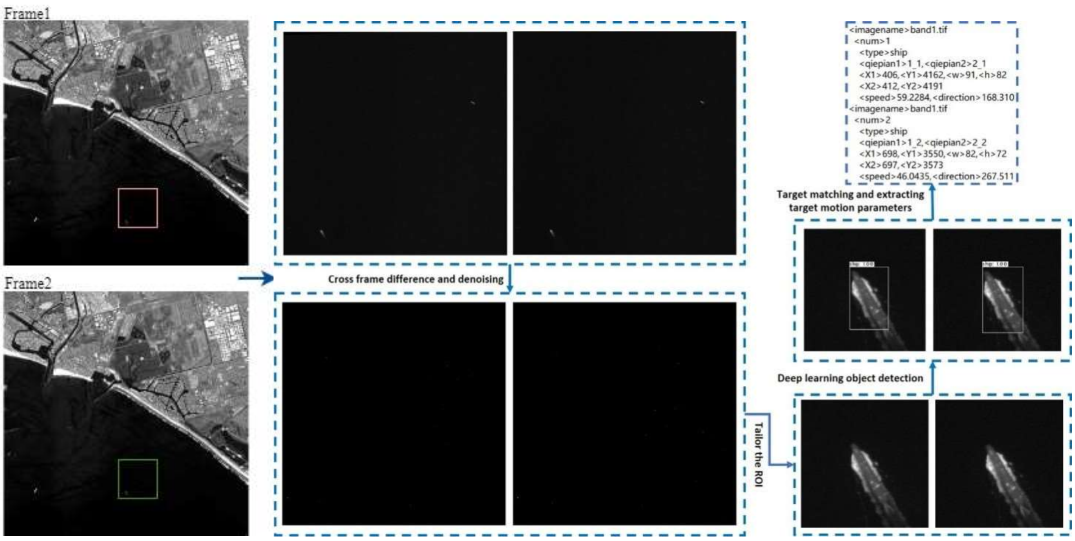

Figure 8. On-orbit processing results of Taijing-IV 01 satellite, including the original images acquired by the satellite, pre-processing results, frame difference results, cropping and target detection results, target matching, and motion parameter estimation results.

Based on calculations, the speed of the vessel is approximately 16 m/s, consistent with the typical cruising speed of a speedboat. The direction, clockwise from true north, forms an angle of 312 degrees. Its position is at approximately 118.116 degrees west longitude and 33.692 degrees north latitude. It is speculated that the vessel is heading towards the nearby port of Los Angeles for docking.

5. Conclusion

This paper's proposed on-orbit moving target detection and correction method improve target detection accuracy and efficiency under satellite energy limitations. In this approach, there are positive interactions between the different steps; YOLO target detection reduces the false alarm rate, while the ROI cropping improves the efficiency of deep learning target detection. The extraction of regions of interest (ROIs) from the motion target area significantly reduces the amount of detection data, with reductions of 63% and 89% compared to grid slicing

https://doi.org/10.5194/isprs-annals-X-4-2024-203-2024 | © Author(s) 2024. CC BY 4.0 License.

and centre cropping methods, respectively. Through enhancements to the YOLOv8 network, we significantly improve the detection capability of small targets in satellite images. Notably, the model size is reduced by 79%, while the mean Average Precision (mAP) increases by approximately 1.8% and 4.5%, and the detection speed rises by 26%.

The satellite automatically detects the targets on-orbit and acquires slice images of the target, which makes the transfer of entire stripe data unnecessary and dramatically improves the overall efficiency of the satellite remote sensing information application. However, there are still some shortcomings in the whole research work, and the problem of incomplete boundaries still exists. In our future work, we will continue to improve the algorithm to balance efficiency and performance as satellite equipment. Moreover, to solve the problem of incomplete target boundary and improve the accuracy of target matching. Ultimately, the satellite will provide more data that will allow us to improve target detection accuracy by labeling more targets, such as ships, aircraft, and other moving targets. It will also provide future researchers with a richer moving target data set.

References

Cao, Y., Wang, G., Yan, D., & Zhao, Z., 2015. Two algorithms for the detection and tracking of moving vehicle targets in aerial infrared image sequences. *Remote Sensing*, 8(1), 28.

Data and Application Centre for High Resolution Earth Observation System in Hubei, 2019. Data and application service of Luojia-1 01 satellite. *Satellite Application*, (5), 26-29.

Fiaz, M., Mahmood, A., Javed, S., & Jung, S. K., 2019. Handcrafted and deep trackers: Recent visual object tracking approaches and trends. *ACM Computing Surveys (CSUR)*, 52(2), 1-44.

Kang, J. Z., Wang, G. Z., He, G. J., Wang, H. H., Yin, R., Jiang, W., & Zhang, Z., 2020. Moving vehicle detection for remote sensing satellite video. J. *Remote Sens*, 24, 1099-1107.

Liang, H., Wang, Q., Zhang, Q, Li, C., 2021. Small object detection technology: a review. *Computer Engineering and Applications*, 57(01), 17-28.

Liu, X. N., 2016. Key technologies in on-board real-time imaging processing for spaceborne SAR. Beijing Institute of Technology, Beijing.

Liu, Y., Zhang, H., Yang, J., et al, 2021., A new algorithm of moving target tracking and detection based on optical flow techniques with frame difference methods. *Electronic Design Engineering*.

Liu, Z., Chen, Y., Cheng, J., 2021. Scale adaptive kernel correlation filter tracking combined with frame difference method. *Computer Technology and Development*, 31(2), 5.

Luo, Y., Liang, Y., Ke, R., & Luo, X., 2018. Traffic flow parameter estimation from satellite video data based on optical flow (No. 18-02345).

Ming, D. P., Luo, J. C., Shen, Z. F., Wang, M., & Sehng, H., 2005. Research on information extraction and target recognition from high resolution remote sensing image. *Cehui Kexue/ Science of Surveying and Mapping*, 30(3), 18-20.

Ou, X. F., Yan, P. C., & Wang, H. P., 2020. Research of moving object detection based on deep frame difference convolution neural network. *Acta Electron. Sin*, 48(12), 2384-2393.

Qi, L., 2020. Research of ship detection in SAR images based on depth features. University of Electronic Science and Technology of China.

Ren, S., He, K., Girshick, R., & Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28.

Sharma, G., Merry, C. J., Goel, P., & McCord, M., 2006. Vehicle detection in 1 - m resolution satellite and airborne imagery. *International Journal of Remote Sensing*, 27(4), 779-797.

Wang, N., Shi, J., Yeung, D. Y., & Jia, J., 2015. Understanding and diagnosing visual tracking systems. *IEEE International Conference on Computer Vision*, (pp. 3101-3109).

Yang, C., Zhou, C., 2019. Review of moving target detection algorithms for UAV video images. *Chinese Journal of Liquid Crystals and Displays*, 34(1), 98-109.

Yu, Y., Zheng, H., 2006. Vehicle detection from high resolution satellite imagery based on the morphological neural network. *Journal of Harbin Engineering University*, 27(s1), 189-193.

Yuan, Y., He, G., Wang, G., Jiang, W., Kang, J. 2018. Remote sensing satellite video vehicle detection method based on fusion of background difference and inter-frame difference. *Journal of University of Chinese Academy of Sciences*, 1, 50-58.

Yue, Z., Qin, Z., Li, J., Shi, Y., 2020. Design of in-orbit processing mechanism for space-earth integrated information network. *Journal of CAEIT*, 15(06), 580-585.

Yu, Y., 2016, Multiple objects real-time tracking and detection technology in satellite video based on multi-core DSP. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences.

Yu, Y., Zhang, T., Guo, L., He, X., 2017. Moving objects detection on satellite video. *Chinese Journal of Liquid Crystals and Displays*, 32(02), 138-143.

Zhang, Z., Yin, T., Wu, H., 2018. Comparison and analysis of moving target detection methods for video star data. *Geomatics and Spatial Information Technology*, 41(3), 3.