Video SAR Moving Target Detection Based on a YOLO Framwork

Qingfeng Tan[1,a], Anxi Yu[1,b,*] , Zhihua He[1,c] and Jiahao Tian[1,d]

[1]*College of Electronic Science and Technology National University of Defense Technology Changsha, 410073, China*

a. 17678313758@163.com, b. yu_anxi@nudt.edu.cn, c. hezhihua@nudt.edu.cn, d. tianjiahao@nudt.edu.cn.

[*]*Anxi Yu*

Keywords: Video synthetic aperture radar, moving target detection, deep convolutional networks, YOLO framework.

Abstract: Video Synthetic Aperture Radar (SAR) is a land imaging mode where radar operates in spotlight mode for a long period of time. Its high frame rate imaging feature can provide the position of the target of interest in the spotlight scene for real-time detection and tracking. The shadow features of video SAR moving targets in the video sequence are difficult to be extracted using traditional methods such as frame difference method and background difference method. At the same time, due to the complex processing steps of traditional method, the detection threshold of targets is usually dynamically adjusted to detect different states of targets. Detection effectiveness and robustness are difficult to meet the requirements of moving target detection applications. However, due to the high number of feature extraction networks in deep convolutional networks, multi-scale shadow features of moving targets can be effectively extracted after convolution operations, and because the deep learning method converts the target detection problem into a regression problem, the detection speed much higher than traditional methods. In this paper, we propose a state-of-the-art moving target detection method based on a YOLO framework. First, data augmentation is performed on the Sandia ViSAR data. At this stage, geometric transformation and color transformation methods are used. Second, in the stage of training the detection model, K-Means clustering is used to screen the sample target by region of interest, determine the YOLO anchor, and finally perform target detection on the testset sample, and the detected AP as an indicator is given for evaluating the detection method process, and do a comparison experiment with the traditional video SAR moving target detection method to verify the effectiveness of this method. The experimental results show that the video SAR moving target detection method based on the YOLO framework increases the detection probability by 28% compared with the traditional frame difference method.

1. Introduction

Traditional synthetic aperture radar (SAR) systems can provide reliable high-resolution imaging capabilities and can operate through clouds for long periods of time in all weather. With the extension of the synthetic aperture time, the SAR system can provide higher azimuth resolution, but the frame rate of traditional SAR system imaging is far lower than the need for real-time target detection and tracking. The concept of video SAR was first proposed by Sandia National Laboratory. Its goal is to simultaneously perceive moving and static targets in the scene. Initially, limited by the device level at the time, the concept did not receive much attention. Until 2012, SNL published four video SAR imaging results on its official website, and its superior system performance has attracted more and more attention. These four videos are (1) SAR imaging result of moving target scene video; (2) SAR imaging result of stationary target scene video; (3) Speed independent continuous tracking result; (4) Video SAR system and infrared system imaging comparison result. These four video results respectively reflect the capabilities of video SAR: high frame rate, high resolution, moving target detection, and cloud penetration and fog penetration. Ground moving targets show defocus characteristics in the video SAR sequential images, which usually deviate from the actual position of the target. It is difficult to achieve target detection and tracking through the target image, but there is no echo reflection in the target area and its shadow area. It shows obvious shadow features in the video SAR sequential image. There is no distortion such as defocusing and offset in the video SAR sequential image. Therefore, the moving target shadow feature can be used to achieve robust moving target detection and tracking processing, with high positioning accuracy, The advantages of high detection probability and no minimum detectable speed limit. However, compared with the characteristics of moving targets in optical video, SAR image shadows are related to factors such as background scattering characteristics, radar equivalent noise coefficient, independent frame rate and geometric resolution, target size and movement speed, and are likely to cause loss of SNR in shadow areas. Thereby significantly reducing detection performance.

In this paper, we present a novel video SAR moving target detection method based on a YOLO framework. First, data augmentation is performed on the Sandia ViSAR data. At this stage, random erasing and mix-up methods are used. Second, in the stage of training the detection model, K-Means clustering is used to screen the sample target by region of interest, determine the YOLO anchor, and finally perform target detection on the testset sample, and the detected AP as an indicator is given for evaluating the detection method process, and do a comparison experiment with the traditional video SAR moving target detection method to verify the effectiveness of this method.

2. Problem Formulation

2.1. Shadow Formation Mechanism of Moving Target

The formation mechanism of shadows in a single frame of video SAR image is shown in Figure. 1. The left picture is a top view, and the right picture is a side view. There are 2 targets in the left picture. Target 1 on the left is a stationary target,marked with a solid green frame, and target 2 on the right is a moving target, marked with a green dashed frame to mark its beginning and end positions, and the shaded area is represented by a red bordered gray frame. Target 2 moves to the lower right at a speed of v_t, and the angle between it and the azimuth is θ. The dimensions of the two targets are both $L_t \times W_t \times H_t$, the viewing angle under the radar is α, the synthetic aperture time is T, and the scene imaging geometry is shown in the coordinate axis in the Figure. 1.

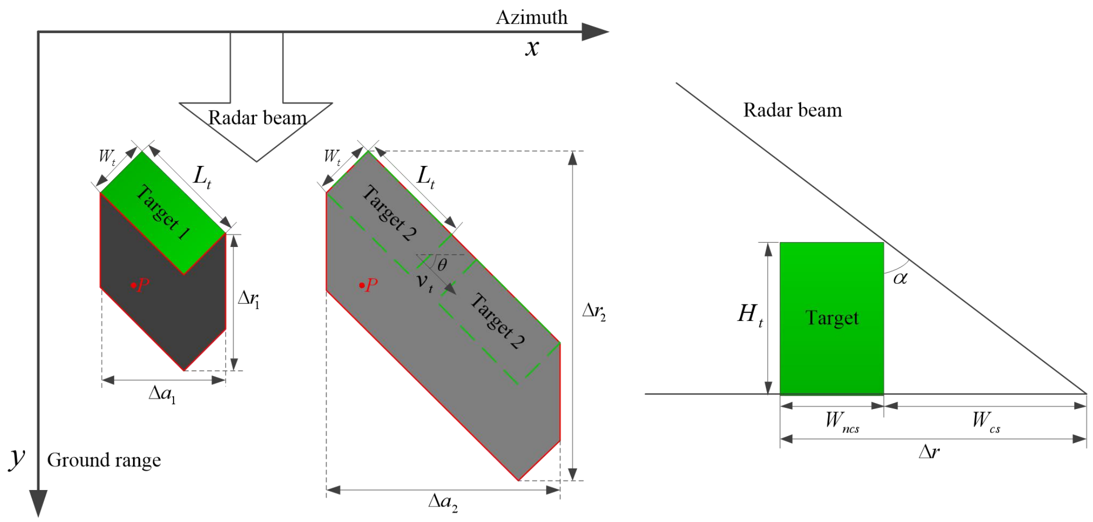


Figure 1: Shadow formation mechanism of moving target.

Analyzing the geometric relationship in Figure. 1, we can see that the shadow area of target 1 is composed of its projection shadow, and its azimuth length Δa_s satisfies

$$\Delta a_s = L_t \cdot \cos\theta + W_t \cdot \sin\theta \tag{1-1}$$

Distance length Δr_s is

$$\Delta r_s = r_{cs} = H_t \cdot \tan\theta \tag{1-2}$$

Where r_{cs} represents the distance length of the projected shadow of the target.

When the target moves, its energy will shift in the azimuth and leave a shadow on the real position, which is called a non-projected shadow. The target 2 shadow area is the combination of its projected shadow and target shadow, and its azimuth length Δa_m can be expressed as

$$\Delta a_m = L_t \cdot \cos\theta + W_t \cdot \sin\theta + v_t \cdot T \cdot \cos\theta \tag{1-3}$$

$v_t \cdot T \cdot \cos\theta$ represents the expansion of the shadow area caused by the target movement.

2.2. Traditional Frame Difference Method and Background Difference Method Processing Flow

The processing flow of traditional frame difference method and background difference method is shown in Figure. 2.

The background difference method is a traditional method of video moving target detection. It extracts the moving area by thresholding the difference between the current frame and the background template. This method is relatively simple. When the background is static, the noise is smooth, the target is clear and the size is large, the method has a better effect on moving target extraction.

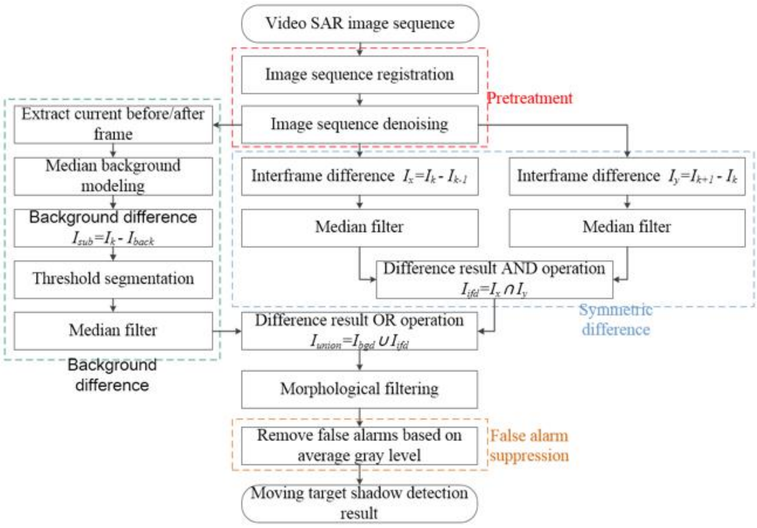


Figure 2: Traditional method process.

Symmetrical difference means that the current frame and its two frames before and after are inter-frame differences respectively, and the obtained difference results are intersected to obtain the contour information of the moving target. The difference method between two adjacent frames will cause more moving targets to be detected than in reality, and only compares and extracts the difference between the two frames, ignoring the information of the overlapping part of itself.

Regardless of whether it is the background difference method or the inter-frame difference method, there is a key problem that the algorithm has high time complexity and cannot be applied to the practical application of video SAR.

2.3. Training YOLO Framework Process

Training the video SAR moving target detection model of the YOLO framework is mainly divided into sample data preprocessing and model training. The training method process is shown in Figure. 3.

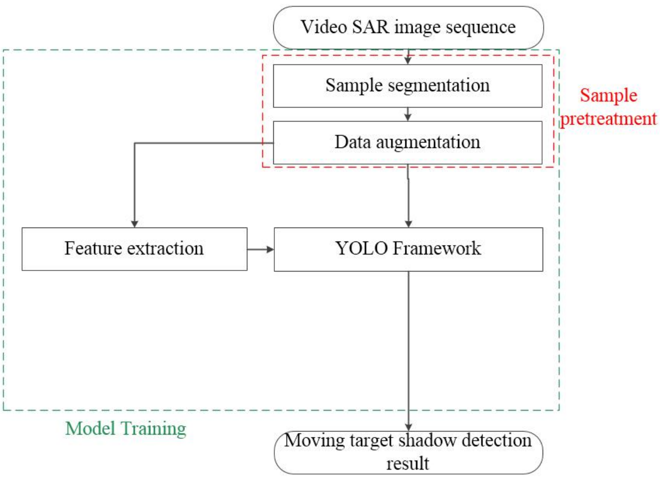


Figure 3: Training YOLO framework process.

Before training the video SAR moving target detection model, the training data needs to be preprocessed. The selection of the anchor point of the YOLO network needs to roughly match the appearance position of the moving target in the training data. At this stage, the video SAR moving target detection method based on the YOLOV3 uses K-Means clustering to screen the sample targets and determine the YOLO anchor points. After determining the YOLO anchor point, the YOLO framework can increase the AP of training and detecting moving targets in video SAR samples by nearly 3 times after initial training 100 Epoch, as shown in Table. 1.

Table 1: Cluster analysis.

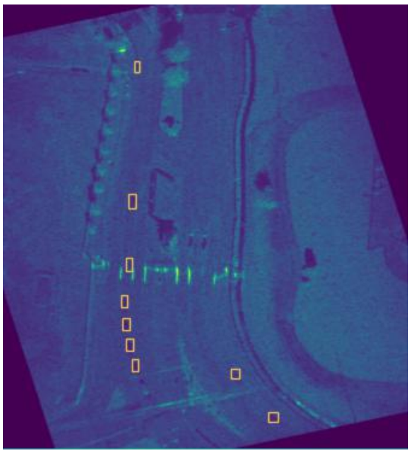
	No clustering algorithm used	K-Means clustering method is used
AP	0.19	0.63

In order to improve the generalization performance of the detection method, this paper uses data enhancement methods such as geometric transformation method and color transformation method. The geometric transformation method uses various operations such as rotating the training sample by multiple angles, horizontal flipping, cropping, deformation, and scaling. Under the premise that the detection of a single frame of video SAR does not consider the direction of the moving target, it can relieve the sample data to some extent The generalization performance of the detection method caused by insufficient. Color transformation achieves the purpose of expanding the sample by changing the sequence of the color channels of the sample.

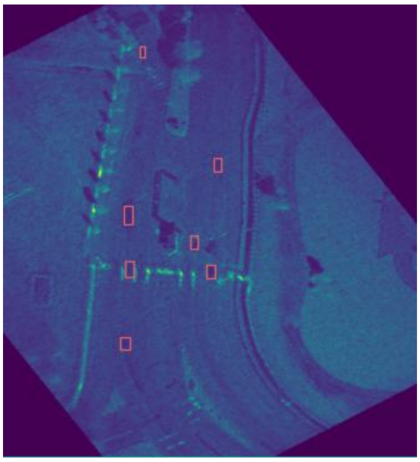
3. Experiment and Analysis

In this section, the training detection model first selects the first 300 frames of Sandia SAR video as the training set test set segmentation, 210 frames are used as the training set, and 90 frames are used as the test set. After 300 Epoch training, the loss gradually converges to about 0.7, and the AP is in IoU. When the sum confidence threshold is given as 0.3 and 0.9, it reaches 0.79.

Finally, the detection results of the video SAR moving target detection method based on the YOLO framework are given, as shown in Figure. 4.



(a) Frame 108.



(b) Frame 264.

Figure 4: The detection results of the video SAR moving target.

In Figure. 4, (a) is a frame extracted from the training set, and (b) is a frame extracted from the test set. Due to space limitations, this article shows two frames in which all moving target shadows are successfully detected. Finally, the detection probability is defined as the ratio of the number of correct moving targets detected in each frame to the number of real moving targets in each frame. Extract the test results in Matlab simulation. The detection probability of the video SAR moving target detection method based on the YOLO framework is obtained, as shown in Figure. 5.

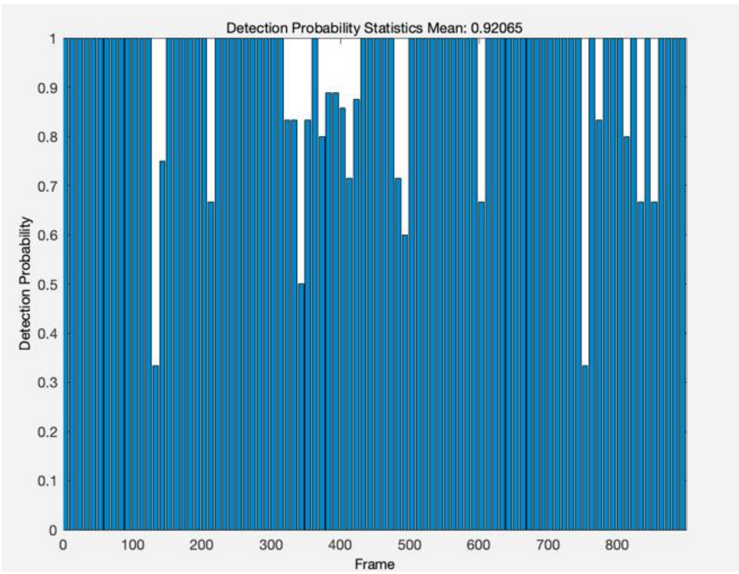


Figure 5: The detection probability of YOLO framework.

Finally, the detection results of the traditional frame difference method and background difference method are shown in Figure. 6.

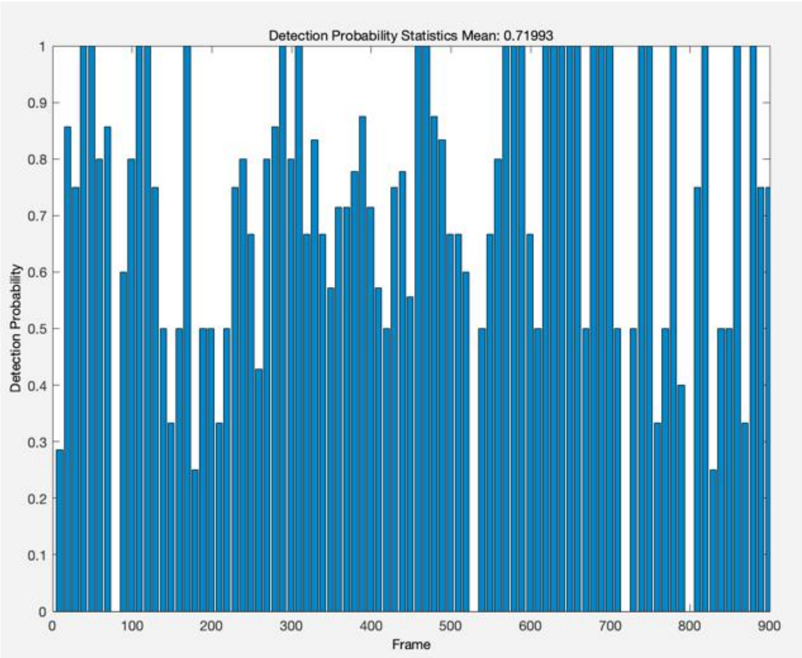


Figure 6: The detection probability of traditional method.

4. Conclusions

The experimental results in Figure. 5 and Figure. 6 show that the video SAR moving target detection method based on the YOLO framework increases the detection probability by 28% compared with the traditional frame difference method. And compared to traditional methods for processing video SAR moving targets, the timeliness of the video SAR moving target detection method based on the YOLO framework is greatly improved. In the simulation experiment, the former needs to be processed for 10 minutes while the latter is training It can be processed within 1 minute after completion, which is of great research value.

References

[1] Li, Zihan, et al. "An Enhanced V-BM3D Algorithm for VideoSAR Denoising Combined with Temporal Information." 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP). IEEE, 2019.

[2] Li, Zihan, et al. "A Robust Image Sequence Registration Algorithm for Videosar Combining Surf with Inter-Frame Processing."IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019.

[3] Ding, Jinshan, et al. "Video SAR moving target indication using deep neural network." IEEE Transactions on Geoscience and Remote Sensing 58.10 (2020): 7194-7204.

[4] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

[5] Taylor, Luke, and Geoff Nitschke. "Improving deep learning using generic data augmentation." arXiv preprint arXiv:1708.06020 (2017).

[6] Zoph, Barret, et al. "Learning data augmentation strategies for object detection." European Conference on Computer Vision. Springer, Cham, 2020.