Fast Segmentation of Solar Panels in Satellite ISAR Images Using a Multitask-YOLO Network

YAO Yuqing[1], *WANG Ling*[1*], *WANG Lianzi*[1], *ZHANG Gong*[1], *WU Bin*[2], *ZHU Daiyin*[1]

1. Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P. R. China;
2. Aerospace System Engineering Shanghai, Shanghai 201109, P. R. China

(Received 28 July 2023; revised 8 December 2023; accepted 20 March 2024)

Abstract: With the rapid development of space technology, the situation awareness ability of spacecraft is increased. As compared to the optical sensors, inverse synthetic aperture radars (ISARs) have the capability of high-resolution imaging in all day from far range regardless of the light condition. Furthermore, the component recognition is much desired by the accurate evaluation of the threat degree of surrounding spacecrafts. In this paper, we propose a multitask-you only look once (Multitask-YOLO) network based on the YOLOv5 structure for recognition and segmentation of solar panels of satellite ISAR images. Firstly, we add a segmentation decoupling head to introduce the function of segmentation. Then, the original structure is replaced with spatial pyramid pooling fast (SPPF) to avoid image distortion, and with distance intersection over union (DIoU) to speed up convergence. The accuracy of segmentation and recognition is improved by introducing an attention mechanism in the channels. We perform the experiments using simulated satellite ISAR images. The results show that the proposed Multitask-YOLO network achieves efficient and accurate component recognition and segmentation. As compared to typical recognition and segmentation networks, the proposed network exhibits an approximate 5% improvement in mean average precision (mAP) and mean intersection over union (mIoU). Moreover, it operates at a higher speed of 16.4 GFLOP, surpassing the performance of traditional multitask networks.

Key words: Multitask-YOLO; space objects; inverse synthetic aperture radar (ISAR) images; target recognition and segmentation

CLC number: TN957.52 **Document code:** A **Article ID:** 1005-1120(2024)02-0253-10

0 Introduction

With the rapid development of space technology, the number of spacecrafts is increasing. The orbital service technology must observe the target when performing operations such as capture, docking, and maintenance. The observation results determine whether the final operation can be successful. Therefore, it is necessary to study recognition technology to monitor the attitude of space targets. Current space sensors rely on optical technology, these devices are limited by observation periods. Inverse synthetic aperture radars (ISARs) have the advantage of providing high-resolution imaging all day[1]. These unique properties make ISAR an important instrument for ensuring the safety of spacecraft. From the ISAR images, the feature information of the target such as shape, size, and attributes can be extracted[2]. With the development of radar technology, the resolution of ISAR imaging for space targets has been greatly improved[3]. The features of the target are more clearly on ISAR image[4]. Therefore, the research into technology for recognition and segmentation of space targets based

*Corresponding author, E-mail address: tulip_wling@nuaa.edu.cn.

How to cite this article: YAO Yuqing, WANG Ling, WANG Lianzi, et al. Fast segmentation of solar panels in satellite ISAR images using a Multitask-YOLO network[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2024, 41(2): 253-262.

http://dx.doi.org/10.16356/j.1005-1120.2024.02.010

on ISAR images is more necessary[5].

Compared to optical images, radar images usually have lower resolution and less defined edges. These characteristics make it difficult to recognize the details of space targets[6]. Therefore, it is important to study a suitable algorithm for recognition and segmentation of ISAR images of space objects[7].

At present, most algorithms based on geometric features are used for target recognition of ISAR images. Yang et al.[8] used trace transform in the longest axis region of the space object and generated the trace matrix of the ISAR image as feature vectors to recognize the target. Ji et al.[9] extracted the image features for space target recognition by two-dimensional wavelet transform. Ju et al.[10] proposed a novel ISAR image segmentation method using a mixed multiscale autoregression model with spatial variation to segment ISAR images. However, due to the characteristics of ISAR images, these algorithms are no longer suitable for component recognition and segmentation. In traditional machine learning algorithms, feature extraction rules are often created manually, and such methods are not representative and cannot recognize the differences between similar categories. The ability of deep learning to learn features by itself can handle this problem. The features of the original image can be better extracted by layer-by-layer transformation. Pei et al.[11] proposed a parallel network topology with multiple inputs for SAR image recognition. Li et al.[12] designed a network that can recognize SAR ships by interactively updating the parameters of sub-networks. Deep learning is currently used in target recognition of SAR images[13]. Aiming at the problem that the traditional method used in the segmentation of ISAR images does not achieve the ideal result. Zhu et al.[14] used U-Net and Siamese network to complete the binary semantic segmentation and binary mask matching of ISAR images and assign the MASK to the corresponding ISAR component. Kou et al.[15] proposed a novel semantic segmentation method for ISAR images of space targets to overcome the relatively weak problem of the supervised learning model. Deep learning is gaining increasing attention in ISAR images[16], but there is still a lack of application for target component segmentation, especially for ISAR images. Therefore, it is necessary to develop a suitable algorithm for segmenting ISAR images into target components.

To improve the segmentation of components of space objects in ISAR images, the multitask-you only look once (Multitask-YOLO) network is proposed based on the YOLO framework, and a segmentation decoupling head is added. At the same time, the original structure is replaced with spatial pyramid pooling fast (SPPF) to avoid image distortion and distance intersection over union (DIoU) to accelerate convergence. Moreover, the proposed network is faster than classical recognition and segmentation networks.

1 Multitask-YOLO Network Architecture

YOLO is an efficient deep neural network for object recognition and localization. YOLOv5[17] inherits the advantages of YOLO's predecessor while improving optimization and recognition accuracy, being less complex, and faster than other models. Therefore, we choose YOLOv5 as the base model.

The proposed Multitask-YOLO network consists of input, backbone network and neck network for feature extraction, and head output layer for output. The specific framework is shown in Fig. 1. The improved network mainly consists of the following functional modules: (1) Focus module—Reducing the amount of computation; (2) cross-stage backbone learning (CBL) module—Ensuring accuracy and reducing the amount of computation; (3) bottleneck with cross stage partial network (BottleneckCSP) module—Enhancing the capability of nonlinear modeling and feature representation in deep networks; (4) SPPF module—Facilitating feature fusion.

(1) Input: In the training phase of the network, the dataset of space objects is extended by image mosaic operation. The algorithm combines four images by random scaling, cropping, and ordering. This method can combine multiple images into one, which improves the training speed of the network and enriches the dataset.

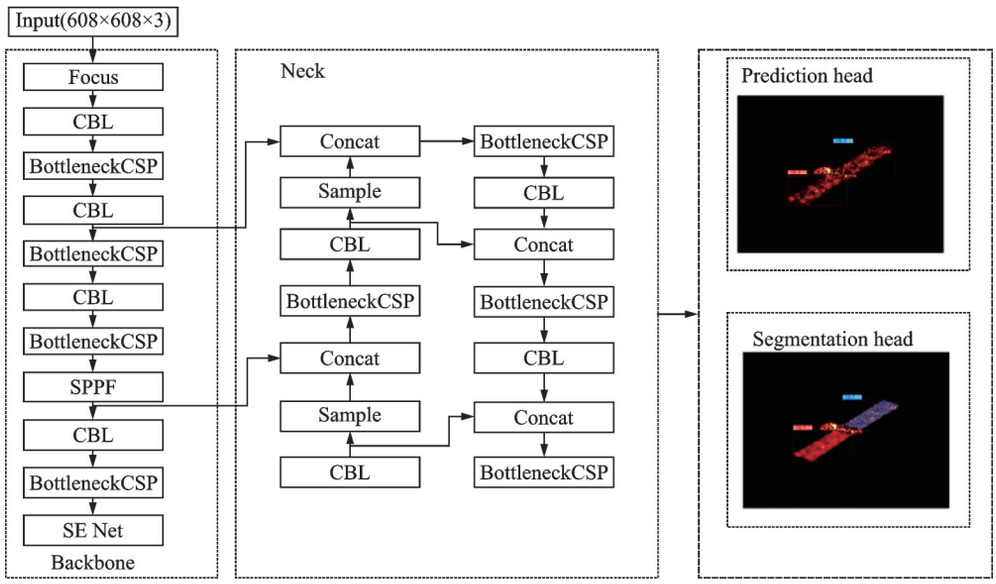

Fig.1 The proposed Multitask-YOLO network structure

(2) Benchmark networks (Backbone): Benchmark networks are high-performance classifiers for extracting general features. The focus structure is used in the improved network, which is used to scale the size of the input image. Through this structure, attention is directed from a single image to multiple images, so that the speed of model inference can be improved. The main idea is to crop the input image by slicing operation. The original size of input image is 608×608×3, after slicing and concatenation, a feature map with a size of 304×304×12 is output. After the map is put into the Concat layer with 32 channels for the convolution operation, a feature map with a size of 304×304×32 is output.

The squeeze-and-excitation network (SE Net) attention mechanism[18] is added to the original backbone to improve the accuracy. The attention of each channel in the feature map is determined. The attention weight is assigned to each feature channel, so that the attention is directed to the useful channels of the feature map for the current task, and the useless feature channels for the current task are suppressed.

SE Net can learn the correlation between the channels and draw the attention to the channels. Fig.2 shows the schematic diagram of SE Net model, which can improve the accuracy of part recognition.

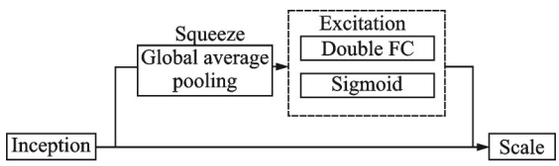

Fig.2 Schematic diagram of SE Net structure

First, we should input the feature map with dimensions $H\times W\times C$ into the global average pooling. Then the global average pooling of the input is squeezed so that the spatial features are compressed to obtain a $1\times 1\times C$ feature map. At the same time, a weight is generated for each feature channel. The channel feature learns the compressed feature graph to obtain with channel attention. Finally, the feature map is multiplied by the normalized weight coefficients obtained channel by channel and then combined with the original input feature map to produce the feature map with channel attention. The method is based on the following three operations to rescale the previous features.

① Squeeze: The compression of features is performed along the spatial dimensions. After the image is put to global average pooling, the operation transforms each two-dimensional feature channel into a real number with a global perceptual field. The output dimensions must match the number of feature channels. This symbolizes the global distribution of the responses to the feature channels and makes the global perception available to the layers near the input.

② Excitation: This step is a mechanism similar to gates in recurrent neural networks. Weights are generated for each feature channel by selecting sigmoid as the activation function after double fully connected (FC) layers. In this way, the parameters learn the correlation between feature channels.

③ Scale: The weights of Excitation's output are considered as the importance of each feature channel after feature selection. The original features are rescaled by multiplying the weights channel by channel with the previous features.

(3) Neck network: The neck network is between the base network and the head network, and can improve the diversity and robustness of features. The radar cross section(RCS) fluctuation caused by changes in perspective leads to the different scattering intensities of each component, which affects the segmentation results. To solve the problem, the sampling structure has been improved in the Neck network, using the SPPF structure combined with feature pyramid network (FPN) and pixel aggregation network (PAN). The effective feature layer obtained in the backbone is fused with the features, so spatial pyramid pooling (SPP) is used with FPN and PAN. Since the radar image has the characteristic of side flap information interference, the redundant features are unnecessary. Therefore, we replace SPP with SPPF for multi-scale feature fusion, so that the characteristics of the target region with strong scattering can be accurately obtained. The structure of SPPF is shown in Fig.3.

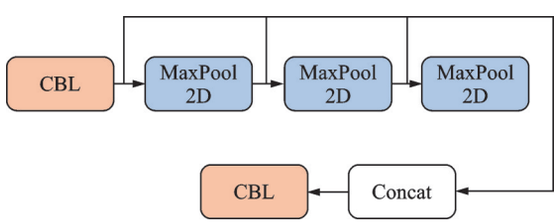

Fig.3 Schematic diagram of SPPF structure

This substitution can improve the calculation speed. After the FPN part is sub-sampled, the PAN is up-sampled to achieve feature fusion, so the detailed and semantic information from different scales can be effectively fused, the region features with weak scattering can be learned more accurately, and the features are not split too finely. The structure of the sampling is shown in Fig.4.

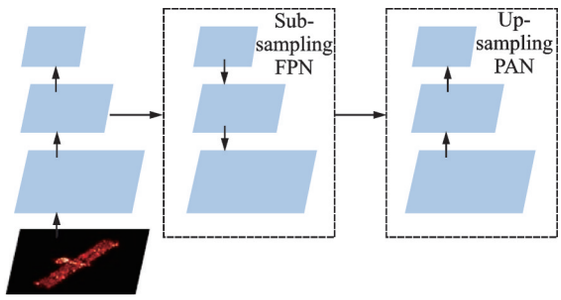

Fig.4 Schematic diagram of sampling structure

(4) Head output layer: The head output layer is responsible for finalizing the output of the component recognition results of the spacecraft ISAR image. The total loss function consists of classification loss, localization loss, and confidence loss. The overall loss is the sum of the three. To mitigate the decrease in speed, the training method incorporates the semantic segmentation loss. The DIoU is used instead of the Smooth L1 Loss function to improve the realism of prediction and the convergence speed. The calculation formula is shown as

$$\mathrm{DIoU}=\mathrm{IoU}-\frac{\rho^2(b,b^{\mathrm{gt}})}{c^2}=\mathrm{IoU}-\frac{d^2}{c^2} \tag{1}$$

$$-1\leqslant\mathrm{DIoU}\leqslant 1 \tag{2}$$

where b represents the parameter for the predicted center coordinates, which is the center point of the red box; b^{gt} represents the parameter of the center of the real target bounding box, which is the center point of the blue box; the value of ρ^2 corresponds to the square of the distance between the two center points represented by d; and c represents the length of the diagonal of the minimum outer rectangle of the two rectangles. The specific definitions of the variables are illustrated in Fig.5. There are two rectangular boxes A and B. Shape C is the smallest boxes containing both A and B. If two boxes overlap exactly, $d=0$, $\mathrm{IoU}=1$, and $\mathrm{DIoU}=1-0=1$. If

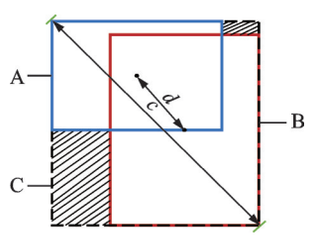

Fig.5 Schematic diagram of DIoU

the two boxes are far apart and d/c converges to 1, IoU = 0, DIoU = 0 − 1 = −1. Therefore, DIoU takes values in the range [−1,1].

Since the segmentation head is added as DeepLabv3+[19] to segment, by utilizing preceding feature layers as its input, the network effectively minimizes computation by establishing parallel connections between channels. The segmented portion improves the characteristics and enhances the accuracy of recognition.

2 Experimental Results and Analysis

As there is no public dataset for radar images of space objects, we utilize the FEKO software to create a database of radar images[19]. Satellites are primarily categorized as reconnaissance satellites, communication satellites, and remote sensing satellites based on their function. The physical structures of each type of satellite are generally comparable. We conducted simulations using three types of satellites as prototypes as shown in Fig.6. Note that the satellites shown in Figs.6(a, d, e) belong to reconnaissance satellites with a size ranging from 0.5 m to 50 m. The satellites in Figs.6(a,b,c) have size over 10 m. These two types of satellites in Figs.6(d,e) have sizes below 10 m.

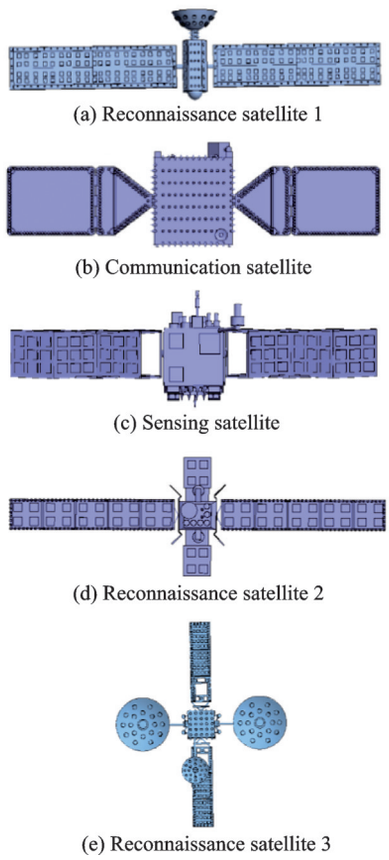

Fig.6 Schematic diagram of FEKO satellite model simulation

We used the FEKO software to simulate the radar images[20]. The simulation parameters are shown in Table 1.

Table 1 Setting of FEKO parameters

Parameter	Start frequency/GHz	End frequency/GHz	Carrier frequency/GHz	Bandwidth/GHz	Number of range cells	Number of pulses
Value	8.35	9.65	9	1.3	100	100

The range resolution is 0.115. The rotation angle accumulated for imaging is 5°, and accordingly the cross-range resolution is 0.191. With this imaging ability, the target with the size larger than 10 m may be imaged appropriately for component recognition. Since we are concerned with component recognition, the satellites larger than 10 m are selected to form the training image set.

We generated the database by imaging with different pitch angles. To obtain effective imaging of the solar panels of the satellites, we select pitch angles of 20°—89°. A simulated model at each pitch angle is continuously observed with the azimuth angle varying from 0° to 360°, where imaging by accumulating every 5° in azimuths. We produced 5 040 images in 20°—89° pitch angles of each mode. In the experiment, we employ the range doppler (RD) algorithm to achieve the images.

After the ISAR images are acquired, we utilized the LabelME software to annotate the image to generate the COCO dataset. The satellite solar panels in the images are labeled by polygons. Following the process of labeling, a JSON file is produced for every image, recording the identified category and

its corresponding spatial coordinates.

The original images have dimensions of 1 680 pixels by 800 pixels. Before input, the images are resized to dimensions of 800×600 to ease training. The training consists of 100 iterations, with a starting learning rate of 0.000 1 and an epoch size of 8. We constructed a database containing 5 040 images of each satellite. The dataset is divided into a training set and a validation set randomly, with the training set consisting of 80% of the database and the validation set consisting of 20% of the database.

Fig.7 shows a comparison of the training loss curve between the Multitask-YOLO network and the original network structure. The horizontal coordinate refers to the length of each step. The proposed network achieves convergence at approximately the 50th epoch, demonstrating faster in convergence and superior performance compared to the original YOLOv5. We evaluate the Multitask-YOLO network by comparing it to several existing algorithms, including YOLOv3[21], BlendMask[22], DeepLabv3+, YOLACT[23], and Mask-RCNN[24].

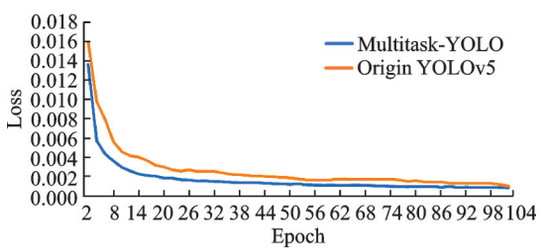

Fig.7 Train loss curves

The ISAR images of satellites are input to the different networks. YOLOv3 is used to recognize the left and right solar panels of the satellite. The panels are segmented using DeepLabv3+, BlendMask, YOLACT, Mask-RCNN, and Multitask-YOLO. Fig.8 displays the results, where left images are sense satellites, middle images are communication satellites and right images are reconnaissance satellites. Fig.8(a) presents the original ISAR images utilized as inputs, and Fig.8(b) presents the images with real annotations. Figs.8(c—h) present the segmentation results of solar panel using DeepLabv3+ (after covering the mask), BlendMask, YOLOv3, Mask-RCNN, YOLACT, and the proposed Multitask-YOLO. The green masks in Figs.8(b—d) represent the left solar panels and the red masks represent the right solar panels.

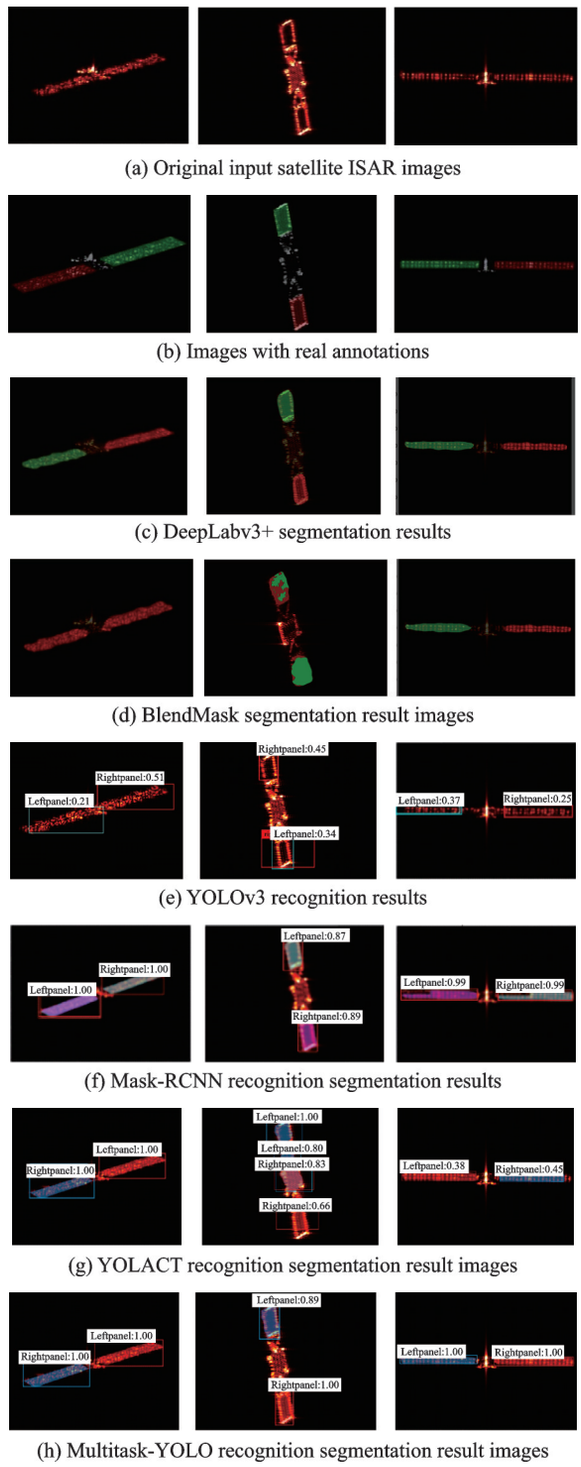

Fig.8 Input images and segmentation results using different networks

Comparing Figs.8(c—h), it is evident that both the DeepLabv3+ and BlendMask networks only focus on performing the segmentation task. According to Fig.8(c), there is a lack of clarity in the connection between the main body and the solar pan-

els, Additionally, due to the similarities in shape, parts of the main body are mistakenly recognized as solar panels. As shown in Fig.8(d), the BlendMask network fails to accurately differentiate between the right and left solar panels when the satellite is rotating, resulting in a low segmentation accuracy. The YOLOv3 network only performs the recognition task. Based on the information presented in Fig.8(e), it is evident that the localization accuracy of the candidate anchors are poor, and the major body cannot be accurately detected and localized. It is because the similarity of color and morphological features between components, leading the network to prioritize speed above accurate calculations. According to Fig. 8 (f), the Mask-RCNN network demonstrates higher accuracy in recognizing and segmenting solar panels. However, it occasionally misclassifies the left and right solar panels. It is due to the lack of judgment on the direction of the main body. The complexity of the network is relatively high and makes the computational time longer. The image in Fig.8(g) shows the output obtained by utilizing the YOLACT network. It can be seen that when the main body shape is similar to that of the solar panels, the method cannot accurately distinguish the target regions. The proposed Multitask-YOLO network can perform both recognition and segmentation tasks simultaneously. The result images are shown in Fig.8(h). The accuracy of separation between the main body and the solar panels has been significantly increased as compared with the previous networks.

Table 2 presents the quantitative evaluation of the results using different networks including mean average precision (mAP), mean intersection over union (mIoU), and running speed. It can be seen that the proposed network, which performs multiple tasks, operates at a speed of 16.4 GFLOP. It is much faster than other networks that perform a single task. In the meanwhile, the recognition and segmentation accuracy surpasses single task networks, exhibiting a 5% enhancement, which outperforms YOLOv3, DeepLabv3+, and Blendmask networks in terms of performance. The multitask-YOLO network has a recognition accuracy that is 0.4% higher and a segmentation accuracy that is 0.3% lower when compared to complicated networks such as Mask-RCNN. The inference speed of Multitask-YOLO is decreased to 1/13 of the inference speed of Mask-RCNN, while maintaining a comparable level of accuracy in recognition and segmentation. YOLACT achieves a speed of 23.4 GFLOP, but its recognition accuracy is approximately 20% lower than the proposed network, and there is a nearly 6% disparity in the segmentation accuracy.

In summary, the Multitask-YOLO network shows the improved performance in solar panels recognition and segmentation compared to a single task network. The capability of both recognition and segmentation is improved. The running speed is significantly greater than that of traditional multitask networks.

An ablation experiment was done on the Multitask-YOLO network to assess the efficiency of each module. Table 3 displays the results from the experiment.

The experimental results in Table 3 demonstrate that incorporating SPPF and DIOU modules enhances both the network's speed and the recogni-

Table 2 Performance comparison of different networks

Network model	mAP (Recognition)/%	mIoU (Segmentation)/%	Operation speed/ GFLOP
DeepLabv3+(Segmentation)		78.59	25.8
BlendMask (Segmentation)		56.32	170.8
YOLOv3(Recognition)	78.04		23.6
Mask-RCNN (Recognition + Segmentation)	86.18	84.03	213.6
YOLACT (Recognition + Segmentation)	63.43	76.44	23.4
Multitask-YOLO (Recognition + Segmentation)	86.58	83.72	16.4

Table 3 Ablation experiment

Network	SPPF	DIoU	SE Net	Seg Head	mAP/%	mIoU/%	Speed/GFLOP
YOLONet 1					79.11		17.4
YOLONet 2	√				80.95		16.6
YOLONet 3	√	√			81.67		16.2
YOLONet 4	√		√		83.61		16.7
YOLONet 5	√	√	√		84.32		16.3
Multitask-YOLO	√	√	√	√	86.58	83.72	16.4

tion accuracy by 1%. The introduction of the SE module increases the recognition accuracy by nearly 4%. YOLONet 4 verifies that the correlation modules can effectively improve the reduction of segmentation accuracy caused by RCS fluctuations. At the same time, upon incorporating the segmentation head, the network acquires the ability to segment, and the recognition accuracy is also improved. Additionally, the network has a reduction in computational complexity by 0.2 GFLOP.

3 Conclusions

Currently, there is no efficient algorithm for recognizing and segmenting ISAR images of space target components. To address this, a Multitask-YOLO network is proposed. This network is capable of simultaneously recognizing and segmenting components. Its effectiveness is tested by simulated ISAR image dataset of space targets. Compared to other single task networks, the proposed Multitask-YOLO network can significantly improve the accuracy of recognition and segmentation. It demonstrates a notable 5% increase in mAP and mIoU, because the result of recognition provides more accurate target location for segmentation. In addition, its speed can reach up to 16.4 GFLOP, surpassing that of traditional multi-task networks. The proposed method is not limited to solar panels and is applicable to the recognition and segmentation of various other components.

References

[1] XUE R, BAI X, CAO X, et al. Spatial temporal ensemble convolution for sequence SAR target classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(2): 1250-1262.

[2] CAO Z, XU L, FENG J. Automatic target recognition with joint sparse representation of heterogeneous multi-view SAR images over a locally adaptive dictionary[J]. Signal Processing, 2016, 126: 27-34.

[3] ZHOU Y, ZHANG L, CAO Y, et al. Attitude estimation and geometry reconstruction of satellite targets based on ISAR image sequence interpretation[J]. IEEE Transactions on Aerospace and Electronic Systems, 2018, 55(4): 1698-1711.

[4] PALADINI R, MARTORELLA M, BERIZZI F. Classification of man-made targets via invariant coherency-matrix eigenvector decomposition of polarimetric SAR/ISAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(8): 3022-3034.

[5] ZHU J, ZHOU J, WU J. Optimization of weighted high-resolution range profile for radar target recognition[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2011, 28(2): 157-162.

[6] MUSMAN S, KERR D, BACHMANN C. Automatic recognition of ISAR ship images[J]. IEEE Transactions on Aerospace and Electronic Systems, 1996, 32(4): 1392-1404.

[7] PARK S H, KIM H T, KIM K T. Segmentation of ISAR images of targets moving in formation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2010, 48(4): 2099-2108.

[8] YANG H, ZHANG Y, XU C. ISAR image space target recognition based on trace features[J]. Journal of Electronics, 2020(3): 431-441.

[9] JI Z, JIN H, ZHANG Y, et al. Space objects recognition based on ISAR image[C]//Proceedings of IEEE International Conference on Communications, Signal Processing, and Systems. Dalian, China: IEEE, 2019: 1215-1222.

[10] JU Y, ZHANG Y, GUO F. ISAR images segmentation based on spatially variant mixture multiscale autoregressive model[C]//Proceedings of IEEE Information Technology, Electronic and Automation Control Conference (IAEAC). Chongqing, China: IEEE, 2018: 2170-2174.

[11] PEI J, HUANG Y, HUO W, et al. SAR automatic

target recognition based on multiview deep learning framework[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 56(4): 2196-2210.

[12] LI D, LIANG Q, LIU H, et al. A novel multidimensional domain deep learning network for SAR ship detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-13.

[13] PARK S H, PARK K K, JUNG J H, et al. ISAR imaging of multiple targets using edge detection and hough transform[J]. Journal of Electromagnetic Waves and Applications, 2008, 22(2/3): 365-373.

[14] ZHU X, ZHANG Y, LU W, et al. An ISAR image component recognition method based on semantic segmentation and mask matching[J]. Sensors, 2023, 23(18): 7955.

[15] KOU P, QIU X, LIU Y, et al. ISAR image segmentation for space target based on contrastive learning and NL-Une[J]. IEEE Geoscience and Remote Sensing Letters, 2023, 20: 1-5.

[16] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[17] JING H, KE Y, ZI J, et al. Solar panel defect detection design based on YOLOv5 algorithm[J]. Heliyon, 2023, 9(8): 18826.

[18] ZHAO H, JIA J, KOLTUN V. Exploring self attention for image recognition[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 10076-10085.

[19] ZHU Runhu, XIN Binjie, DENG Na, et al. Semantic segmentation using DeepLabv3+ model for fabric defect detection[J]. Wuhan University Journal of Natural Sciences, 2022, 27(6): 539-549.

[20] WANG F, EIBERT T F, JIN Y Q. Simulation of ISAR imaging for a space target and reconstruction under sparse sampling via compressed sensing[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(6): 3432-3441.

[21] LIU R, LEN X, LIU Y. A fast image matching algorithm based on YOLOv3[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2021, 38(5): 807-815.

[22] CHEN H, SUN K, TIAN Z, et al. Blendmask: Top-down meets bottom-up for instance segmentation[C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 8573-8581.

[23] BOLYA D, ZHOU C, XIAO F, et al. Yolact: Real-time instance segmentation[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 9157-9166.

[24] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//Proceedings of IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2017: 2961-2969.

Acknowledgements This work was supported in part by the Shanghai Aerospace Science and Technology Innovation Foundation (No.SAST 2021-026), and the Fund of Prospective Layout of Scientific Research for Nanjing University of Aeronautics and Astronautics(NUAA).

Authors Ms. **YAO Yuqing** is working toward a Master degree in information and communication processing in Nanjing University of Aeronautics and Astronautics, Nanjing, China. Her research interests include recognition and segmentation technology on radar imaging and deep learning.

Prof. **WANG Ling** received the B.S. degree in electrical engineering, and the M.S. and Ph.D. degrees in information acquirement and processing from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2000, 2003 and 2006, respectively. In 2003, she joined Nanjing University of Aeronautics and Astronautics, where she is currently a professor in the Department of Information and Communication Engineering. Her current research interests include radar, microwave imaging, image processing and intelligent information processing.

Author contributions Ms. **YAO Yuqing** built the model, realized the method, performed the demonstration and wrote the paper. Prof. **WANG Ling** presented the main idea and gave valuable suggestions to finalize the work. Ms. **WANG Lianzi** provided the training data. Prof. **ZHANG Gong** commented on the paper and approved the submission. Dr. **WU Bin** contributed to the experimental analysis and provided suggestions from the perspective of practical applications. Prof. **ZHU Daiyin** contributed to the discussion and background of the study. All authors commented on the manuscript draft and approved the submission.

Competing interests The authors declare no competing interests.

(Production Editor: ZHANG Huangqun)

基于Multitask-YOLO网络的卫星帆板ISAR图像快速分割

姚雨晴[1]，汪 玲[1]，王莲子[1]，张 弓[1]，吴 斌[2]，朱岱寅[1]

（1.南京航空航天大学电子信息工程学院雷达成像与微波光子技术教育部重点实验室，南京 211106，中国；
2.上海宇航系统工程研究所，上海 201109，中国）

摘要：随着空间技术的飞速发展，空间态势感知能力需求不断增加。与传统光学传感器相比，逆合成孔径雷达（Inverse synthetic aperture radar，ISAR）具有全天候、远距离高分辨率成像的能力，且成像不受光照条件的影响。此外，空间态势感知系统需要对周围航天器进行准确的评估，因此对空间目标部件识别能力的需求日益迫切。本文提出了一种基于YOLOv5结构的Multitask-YOLO网络，用于卫星ISAR图像中卫星帆板的识别和分割。首先，本文添加了分割解耦头来实现网络的分割功能。然后用空间金字塔池快速算法（Spatial pyramid pooling fast，SPPF）和距离交并比算法（Distance intersection over union，DIoU）代替原有结构，避免图像失真，加快收敛速度。通过在通道中引入注意机制，提高了分割和识别的准确性。最后使用模拟卫星的ISAR图像进行实验。结果表明，所提出的Multitask-YOLO网络高效、准确地实现了部件的识别和分割。与其他的识别和分割网络相比，该网络的平均精度（mean Average precision，mAP）和平均交并比（mean Intersection over union，mIoU）提高了约5%。此外，该网络的运行速度高达16.4 GFLOP，优于传统的多任务网络的性能。

关键词：Multitask-YOLO；空间目标；逆合成孔径雷达图像；目标识别与分割