



OPEN ACCESS

EDITED BY

Haiyong Zheng,
Ocean University of China, China

REVIEWED BY

Yangfan Wang,
Ocean University of China, China
Xiaoling Zhang,
University of Electronic Science and
Technology of China, China
Ibrar Ahmad,
University of Peshawar, Pakistan

*CORRESPONDENCE

Lili Zhan

✉ skd992016@sdust.edu.cn

RECEIVED 01 December 2022

ACCEPTED 07 April 2023

PUBLISHED 01 May 2023

CITATION

Yasir M, Zhan L, Liu S, Wan J, Hossain MS,
Isiacik Colak AT, Liu M, Islam QU,
Raza Mehdi S and Yang Q (2023) Instance
segmentation ship detection based on
improved Yolov7 using complex
background SAR images.
Front. Mar. Sci. 10:1113669.
doi: 10.3389/fmars.2023.1113669

COPYRIGHT

© 2023 Yasir, Zhan, Liu, Wan, Hossain, Isiacik
Colak, Liu, Islam, Raza Mehdi and Yang. This
is an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Instance segmentation ship detection based on improved Yolov7 using complex background SAR images

Muhammad Yasir¹, Lili Zhan^{2*}, Shanwei Liu¹, Jianhua Wan¹,
Md Sakaouth Hossain³, Arife Tugsan Isiacik Colak⁴,
Mengge Liu², Qamar Ul Islam⁵, Syed Raza Mehdi⁶
and Qian Yang⁷

¹College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China, ²College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao, China, ³Department of Geological Sciences, Jahangirnagar University, Dhaka, Bangladesh, ⁴National University International Maritime College Oman, Sahar, Oman, ⁵Department of Electrical and Computer Engineering, College of Engineering, Dhofar University, Salalah, Oman, ⁶Department of Marine Engineering, Ocean College, Zhejiang University, Zhoushan, Zhejiang, China, ⁷People's Liberation Army (PLA) Troops No.63629, Beijing, China

It is significant for port ship scheduling and traffic management to be able to obtain more precise location and shape information from ship instance segmentation in SAR pictures. Instance segmentation is more challenging than object identification and semantic segmentation in high-resolution RS images. Predicting class labels and pixel-wise instance masks is the goal of this technique, which is used to locate instances in images. Despite this, there are now just a few methods available for instance segmentation in high-resolution RS data, where a remote-sensing image's complex background makes the task more difficult. This research proposes a unique method for YOLOv7 to improve HR-RS image segmentation one-stage detection. First, we redesigned the structure of the one-stage fast detection network to adapt to the task of ship target segmentation and effectively improve the efficiency of instance segmentation. Secondly, we improve the backbone network structure by adding two feature optimization modules, so that the network can learn more features and have stronger robustness. In addition, we further modify the network feature fusion structure, improve the module acceptance domain to increase the prediction ability of multi-scale targets, and effectively reduce the amount of model calculation. Finally, we carried out extensive validation experiments on the sample segmentation datasets HRSID and SSDD. The experimental comparisons and analyses on the HRSID and SSDD datasets show that our model enhances the predicted instance mask accuracy, enhancing the instance segmentation efficiency of HR-RS images, and encouraging further enhancements in the projected instance mask accuracy. The suggested model is a more precise and efficient segmentation in HR-RS imaging as compared to existing approaches.

KEYWORDS

computer vision, object detection, instance segmentation, HR-RS, YOLOv7, SSDD, HRSID, SAR Complex background images

1 Introduction

SAR is a microwave imaging sensor built on electromagnetic wave scattering properties that may be used in all weather conditions and has some ability to penetrate clouds and the ground. With the ongoing exploitation of maritime resources as well as the increased attention being paid to the monitoring of marine ships, it has special benefits in marine monitoring, mapping, the military, and all of these fields (Li et al., 2022; Liu et al., 2022; Kong et al., 2023; Yasir et al., 2023a; Yasir et al., 2023b). SAR ship detection technique is therefore very important for protecting marine ecosystems, maritime law enforcement, and territorial sea security. Ocean ship monitoring has received a lot of attention (Zhang et al., 2020b; Chen et al., 2021; Xu et al., 2022a; Zhang et al., 2023). Synthetic aperture radar (SAR) is more suited for monitoring ocean ships than optical sensors (Zeng et al., 2021; Zhang and Zhang, 2021a; Xu et al., 2022b; Zhang and Zhang, 2022c) because of its ability to operate in all weather conditions (Zhang and Zhang, 2021b). Ship monitoring is a key maritime task that is crucial for ocean surveillance, national defense security, fisheries management, etc. identification Ship in the SAR picture is a significant area of remote sensing research because it relies on target detection technology, which is in high demand (Wang et al., 2018; Chang et al., 2019; Qian et al., 2020; Su et al., 2022). Ship identification in satellite RS pictures has grown in importance as a research area recently (Nie et al., 2020). The marine transportation sector is now developing extremely quickly. The number of maritime infractions has increased as a result of the quick expansion in ship numbers and shipping volume. Automated ship identification plays an increasingly essential role in maritime surveillance, monitoring, and traffic supervision as well as in the regulation of illegal fishing and freight transit. It can assist in gathering information about ship dispersion. HR-RS pictures are given by a variety of airborne and spaceborne sensors, including Gaofen-3, TerraSAR-X, RADARSAT-2, Ziyuan-3, Sentinel-1, Gaofen-2, and unmanned aerial vehicles (UAV), owing to the quick development of imaging technology in the domain of RS. These HR pictures are being used in the military and the domains of the national economy, such as traffic control, marine management, urban monitoring, and ocean surveillance (Mou and Zhu, 2018; Cui et al., 2019; Su et al., 2019; Sun et al., 2021b). The HR RS pictures are especially well suited for object identification and segmentation in areas like military precision strikes and maritime transportation safety (Su et al., 2019; Wang et al., 2019; Zhang et al., 2020a). Instance segmentation, which may be characterized as a technology that addresses both the issue of object identification and semantic segmentation, has emerged as a significant, sophisticated, and challenging area of research in machine vision. Parallel to semantic segmentation, it has both pixel-level classification and object identification properties, where dissimilar instances must be located even if they belong to the same type (Xu et al., 2021). Since the two-stage object identification algorithm's introduction, other convolutional neural network-based object detection and segmentation methods have appeared, including the R-CNN, Faster R-CNN (Ren et al., 2015), and Mask R-CNN (He et al., 2017).

Deep learning innovation demonstrates inspiring outcomes recently in several fields, including object identification (Zhang et al., 2019a; Zhang et al., 2020c; Zhang et al., 2021a), image classification (Liu et al., 2021b; Zhou et al., 2022a; Zhou et al., 2022b), Segmentation (Liu et al., 2021b; Zhou et al., 2021; Zong and Wan, 2022; Zong and Wang, 2022), and so on (Zhou et al., 2019; Liu et al., 2021a; Wu et al., 2022; Yin et al., 2022; Zhu and Zhao, 2022). Recently, despite the existence of many excellent algorithms, like the path aggregation network (Liu et al., 2018), Mask Score R-CNN (Wang et al., 2020a), Cascade Mask R-CNN (Dai et al., 2016), and segmenting objects by locations (Wang et al., 2020b) and so on (Zhang and Zhang, 2019; Zhang et al., 2019b; Zhang et al., 2021b; Shao et al., 2022; Zhang and Zhang, 2022a; Zhang and Zhang, 2022b; Zhang and Zhang, 2022c; Zhang and Zhang, 2022d), common issues, such as erroneous segmentation edges and the development of global relations, still exist. The extension of the model will lead to dimensional disasters if the long-range dependencies are represented by dilated convolution or by expanding the number of channels. YOLOv7 serves as the basic foundational framework for the development of a framework model for RS picture object identification and instance segmentation in order to get over CNNs' limitations in terms of their capacity to extract spatial information. Detecting and segmenting ships in SAR images is difficult because of the complexity and variety of the images themselves, which include speckle noise, shadows, and cluttered backgrounds. These elements make it challenging to reliably identify ships among other objects in the image and to define the ship's boundaries.

In addition, different from moving targets such as aircraft and vehicles, ship targets often dock side by side near the port, so it is difficult for general detection methods to accurately distinguish each target, resulting in a large number of missing targets. Meanwhile, Ship case segmentation can not only accurately obtain the position of the object, but also effectively achieve the shape information of the target, which can further promote the research of SAR ship recognition. However, at present, a large number of studies only focus on the SAR ship targets detection and do not further achieve the target-level instance segmentation. It is specifically affected by the following factors, (1) the complexity of the instance segmentation model is high, often reaching hundreds of megabytes, which is difficult to be applied. (2) The running efficiency of the instance segmentation algorithm is relatively low, and the initial training of the model takes a long time. (3) There is not enough sample data to train the model, which makes the performance of existing deep learning methods insufficient. In our study, we utilized various data augmentation techniques, such as random flipping, rotation, and scaling, to generate additional samples from the limited dataset. These techniques effectively increase the diversity of the training samples and help prevent overfitting.

To overcome this problem, we propose an improved version of the YOLOv7 object detection algorithm that incorporates an ELAN-Net backbone and feature pyramid network (FPN) to boost the model's capability to extract relevant features from SAR images in complex backgrounds. Our suggested algorithm achieves state-of-the-art effectiveness on two benchmark datasets, demonstrating its effectiveness in addressing the research problem

of accurate ship identification and segmentation in complex SAR pictures. The main contributions in this paper are outlined in the following order:

Λ An upgraded YOLOv7 model has been proposed for instance segmentation ship detection.

Λ An effective feature extraction module has been developed and added to the improved backbone network, enhancing the network's focus on target features and making the process of feature extraction more efficient.

Λ The feature pyramid module is optimized with feature fusion to increase the accuracy of multi-scale target segmentation and further improve the speed of image processing to boost the identification and segmentation performance of the network for multi-scale ship targets.

Λ Two ship datasets, an SSDD dataset, and an HRSID dataset are used to evaluate the efficiency of the suggested technique. To test the model's robustness, two ship datasets are run (which contain images with different scales, resolutions, and scenes).

The paper is structured as follows: Part 2 explains the materials and experimental setup and demonstrates how the study acts as an organizing foundation for the remaining portions of the research. Part 3 provides a description of the research project's results and analyses. It has also shown the model's potential by comparing it with other innovatively made versions. The ablation study is described in Section 4, and Section 5 concludes the paper.

2 Related work

2.1 Deep learning-based instance segmentation

Instance segmentation in SAR photos has the advantage of combining semantic segmentation with object identification. Using semantic segmentation, each pixel of the input picture is separated into logical groups according to where the ship targets are located. It offers a better description and perception of the ship targets because of the more complex interpretation technique. As the first attempt at segmenting CNN, Mask R-CNN (Lin et al., 2017b) adds a mask branch that is analogous to the classification and regression branch in Faster R-CNN in order to forecast the segmentation mask for each region of interest (RoI). Mask Scoring R-CNN (Wang et al., 2020a) utilizes the product of the classification score and the IoU score of the mask to construct the mask score in order to increase the quality of an instance. Cascade Mask R-CNN is created by combining Mask R-CNN and Cascade R-CNN (Chen et al., 2019b). Each cascade framework adds a mask branch to complete the instance segmentation task, combining the best features of the two approaches. In order to improve identification accuracy, Hybrid Task Cascade (Chen et al., 2019b) proposes integrating the concurrent structures of identification and segmentation, which leverage semantic segmentation branches to build a spatial context for the bounding box. In recent years, a number of one-stage algorithms, notably YOLACT (Bolya et al., 2019) and SOLO (Wang et al., 2020b), have appeared that correspond to object identification methods. In addition, a few approaches such as

BlendMask (Chen et al., 2020) and PolarMask (Xie et al., 2020) are built on an item identification network without anchors. Due to their speed benefits, these one-stage techniques are frequently utilized in the domain of autonomous vehicle operation and facial detection. However, in some complex ship identification tasks, the identification technique can only assess a ship's length and contour when they are important details for the particular type of ship. Improvements to the current algorithms for segmenting SAR images by an instance are not currently being made in a substantial way. The HRSID (Lin et al., 2017a) dataset was recently created for the segmentation of ship instances in SAR images.

2.2 SAR images-based ship detection

SAR can continually monitor the planet, in contrast to optical sensors, which are inoperable at night. Because SAR images do not contain information about color, texture, shape, or other aspects, they show ships differently than optical images do. Furthermore, the SAR image has a lot of noise; as a result, identifying SAR images might be difficult for researchers without the appropriate skills. Because there is a dearth of data on tagged SAR ships as an outcome, it is more challenging to identify ships from SAR images. In order to find ships in SAR images, several deep-learning techniques have been used (Sun et al., 2021a; Liu et al., 2022; Sun et al., 2022; Yasir et al., 2022). (Fan et al. 2019b) implemented a multi-level features extractor into the Faster R-CNN for polarimetric SAR ship identification. A dense attention pyramid network was created to identify SAR ships by densely connecting each feature map to the attention convolutional module (Cui et al., 2019). For pixel-by-pixel ship identification in polarimetric SAR photos, a fully convolutional network has been created (Fan et al., 2019a). The feature pyramid structure contained a split convolution block and an embedded spatial attention block (Gao et al., 2019). Against a complex background, the feature pyramid structure can identify ship items with accuracy. Wei et al. (Wei et al., 2020) created a high-resolution feature pyramid structure for ship recognition that combined high-to-low-resolution features. The challenge of ships of various sizes and crowded berthings has been addressed by the development of a multi-scale adaptive recalibration structure (Chen et al., 2019a). A one-stage SAR target identification approach was suggested by Hou et al. (Hou et al., 2019) to address the low confidence of candidates and false positives. (Kang et al. 2017) proposed a method integrating CFAR with faster R-CNN. The object proposals produced by the faster R-CNN used in this method for extracting small objects served as the protective window of the CFAR. Zou et al. (Zou et al., 2020) integrated YOLOv3 with a generative adversarial network with a multi-scale loss term to increase the accuracy of SAR ship recognition. In order to identify and recognize ships in complex-scene SAR images, Xiong et al. (Xiong et al., 2022) suggested a lightweight model that integrated several attention mechanisms into the YOLOv5-n lightweight model.

Results from using CNN methods to identify ships in SAR imagery are impressive. However, there are still two significant

areas of work that need to be addressed. One of these involves methodically combining the most recent advancements in computer vision to connect optical and SAR images. The other seeks to broaden the use of ship identification to further applications, such as instance segmentation. The two SAR image components were combined as part of this study to enhance the images' suitability for RS applications, which is another goal of the investigation.

3 Proposed improved methodology

3.1 Overall structure of our model

In addition to classifying and locating the object of interest in an image, instance segmentation also labels each pixel that is a component of the particular object instance. It enhances the identification process by associating the bounding box and mask with the object. As a result, instance segmentation will help us identify ships more accurately and will also help us deal with crowded sceneries and detect partially occluded ships. Semantic segmentation-based bottom-up and identification-based top-down techniques have been the main focus of case segmentation research for a very long time. The majority of CNN-based models and their derivation models, including RCNN, have been used for computer vision tasks such as object identification, tracking, segmentation, and classification. Faster RCNN (Chen et al., 1993) is improved by a cutting-edge technique known as Mask RCNN (He et al., 2017), which also does instance segmentation using region proposals. Additionally, it locates every instance of the target object down to the pixel level in an image.

YOLO is a single-stage object detector that can forecast a particular object in each area of the feature maps without the aid of the cascaded location classification stage. YOLO categorizes and locates the object using bounding boxes and a particular Convolution Neural Networks (CNN) network. It splits the image into an $S \times S$; $S \in \mathbb{Z}^+$ grid and identifies an object as a grid cell if its

focal point crosses one. A one-stage detection method called YOLO may recognize objects instantly and is very quick (Redmon et al., 2016). The YOLOV7 algorithm, which is now the most sophisticated in the YOLO series, balances the conflict between the quantity of parameters, the amount of calculation, and the performance. It also outperforms earlier iterations of the YOLO series in terms of accuracy and speed. In this paper, we used the improved Yolov7 for segmentation ship detection, and Figure 1 illustrates the outline of the method recommended in the research.

The 1024x1024 SAR images are concurrently supplied to the network feature extraction at the input end, as shown in Figure 1. In order to successfully manage the framework training, the proper ship target labeling must be delivered. The entire deep framework is divided into three sections: the backbone structure, which is primarily used to extract features from the input picture; the feature pyramid, which is used to scale the extracted features and strengthen the expression of the target feature; and the network prediction layer, which predicts the target at three scales. Finally, post-processing techniques like maximum value suppression (NMS) are used to acquire the results of the identification output.

3.2 Improved backbone networks

The two new modules that are added to the backbone structure in this research are as follows: SiLu function is used by the MP-Conv module, the E-ELAN module, and its activation function. The SiLU activation function used by the MP-Conv module is known to be more computationally efficient and effective than the traditional ReLU activation function. By incorporating the SiLU function, the MP-Conv module can better capture relevant features from SAR images, leading to improved object detection performance. Meanwhile, The MP-Conv module adopts the way of double-branch fusion to carry out super downsampling of convolution blocks, which on the one hand improves the operational efficiency of target feature extraction, on the other hand, it can fuse and enhance target feature expression. The E-ELAN module is designed

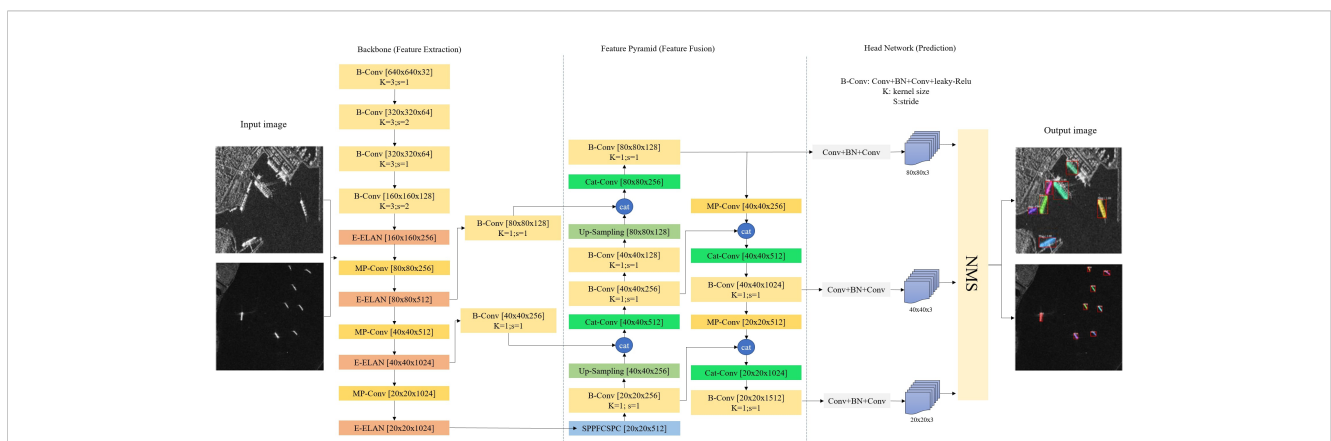


FIGURE 1 The overall structure of the proposed ship detection and segmentation model. E-ELAN, MP-Conv, Cat-Conv, and SPPSPC are some improved modules.

to boost the capability of the algorithm to retrieve spatial information from the SAR image. This is achieved by incorporating an attention mechanism that selectively weighs the feature maps based on their relevance to the final prediction. By selectively weighing the feature maps, the E-ELAN module can help the model focus on the most relevant information, leading to improved detection and segmentation performance. In addition, the E-ELAN module can stack more blocks by considering the shortest gradient path, so as to enhance the feature extraction capability of the network without significantly increasing the complexity of the model.

The E-ELAN module is an effective network structure, as shown in Figure 2A, that enables the network to learn more features and has stronger robustness by managing the shortest and longest gradient routes. The ELAN module has two branches specifically: The first branch involves using a 1×1 convolution to adjust the number of channels. The second branch, which is more difficult, first passes through a 1×1 convolution module to alter the number of channels. Then, run four 3×3 convolution modules to extract features.

The reason for selecting the fourth B-Conv as the branch for channel concatenating in Figure 2 is that we conducted extensive experiments and found that this branch provides the best performance for ship detection. Specifically, we found that by selecting the fourth B-Conv branch, the network can effectively capture features at different scales and resolutions, which is critical for accurate ship instance segmentation detection in complex background SAR images.

Two branches of the MP-Conv (Max-Pooling Convolution) module, as seen in Figure 2B, are employed for downsampling. A Max-pool, or maximal pooling, is used on the first branch. The result of maximizing is downsampling and a 1×1 convolution to change the number of layers. The second branch initially performs a 1×1 convolution to change the number of layers before passing through a convolution block with a 3×3 convolution kernel and a 2 stride. Downsampling is another application for this convolution block. In the end, the two branches' results are combined, the

number of layers equals the number of input layers, but the spatial resolution is decreased by a factor of 2.

In summary, the proposed model structure is designed to enhance the model's ability to extract relevant features from SAR images and to incorporate spatial information through the attention mechanism. These improvements contribute to improved object detection and segmentation performance, as demonstrated in our experiments.

3.3 Improved neck networks

Figure 3 displays the detailed structures of two enhanced modules in the neck network. Figure 3A illustrates how similar the Cat-conv module is to the E-ELAN (Encoder Enhanced Layer Aggregation Network) module, with the exception that it chooses a different number of outputs for the second branch. Three outputs are chosen by the E-ELAN module for final addition, and five channels are chosen by the Cat-conv module for contact. The Cat-conv structure utilized in this article can assist the entire pyramid framework in aggregating multi-scale features, increasing the multi-scale representation of ship targets, which have remarkable multi-scale features in SAR images.

In order to increase the receptive field more efficiently and further promote the algorithm to adapt to different resolution images, we optimize to design of the SPPSPC (Spatial Pyramid Pooling with Spatial Pyramid Convolution) module to replace the original SPP module. As seen in Figure 3B, the first branch has four branches following the Max-pool operation. Through maximal pooling, it obtains various receptive fields. These four distinct branches signify the network's ability to process a variety of objects. That is to say, it has four receptive fields for each of its four separate scales of maximum pooling, which are utilized to differentiate between large and small targets. In this way, the SPPSPC module designed in this paper combines and optimizes the feature reorganization, which can effectively increase the accuracy of the algorithm while greatly reducing the amount of

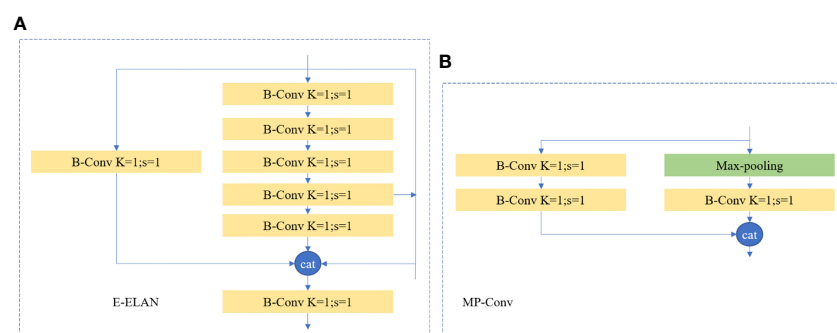


FIGURE 2

The detailed structures of two improved modules in the backbone network. (A) The E-ELAN module. (B) The MP-Conv module. "Conv" means the ordinary convolution-2D layer, "BN" means the batch normalization layer, "Max-Pooling" means the max pooling-2D layer; "K" is the kernel size, and "s" is the sliding step.

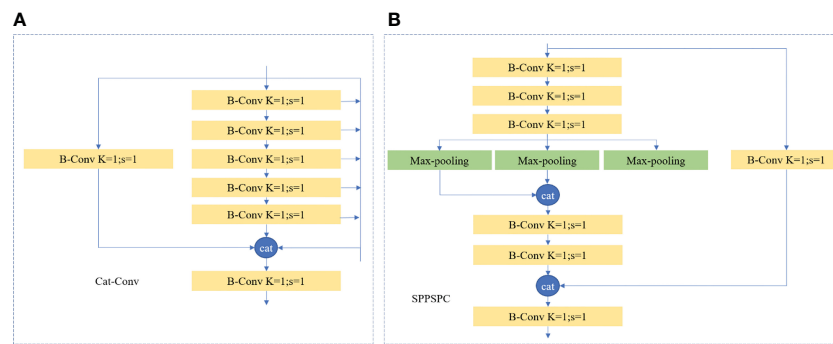


FIGURE 3

The detailed structures of two improved modules in the neck network. (A) The Cat-Conv module. (B) The SPPSPC module. “Conv” means the ordinary convolution-2D layer, “BN” means the batch normalization layer, “Max-Pooling” means the max pooling-2D layer; “k” is the kernel size, and “s” is the sliding step.

computation. The loss function used in our proposed network is a combination of three loss functions: the localization loss, the confidence loss, and the segmentation loss. The localization loss measures the difference between the predicted bounding box and the ground truth bounding box. The confidence loss measures the objectness score and the background score. Finally, the segmentation loss measures the pixel-wise difference between the predicted mask and the ground truth mask. The overall loss function is a weighted sum of these three loss functions, and it is optimized using the stochastic gradient descent (SGD) algorithm.

4 Experimental result and discussions

4.1 Dataset overview

4.1.1 HRSID dataset

The High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation (HRSID) provided by Wei et al. (Lin et al., 2017b) is made up of images from 99 Sentinel-1B imageries, 36 TerraSAR-X, and 1 TanDEM-X imagery. The resolutions of the 800 x 800-pixel images, which contain 16951 ships and 5604 sliced SAR images, range from 1 to 15 meters.

4.1.2 SSDD dataset

The first and most important stage in applying deep learning algorithms to recognize ships is the construction of a substantial and comprehensive dataset. As a result, the experiment makes use of the SSDD (Li et al., 2017) dataset, which contains 1160 SAR

pictures from Sentinel-1 TerraSAR-X, and RadarSat-2 with resolutions ranging from 1m to 15m and polarizations in HV, HH, VH, and VV (Table 1). Scenes of offshore ships and inshore ships are both present in the collection as background elements.

4.2 Implementation setting

The experiments are all run on an Intel Core i9-9900KF CPU and an NVIDIA Geforce GTX 2080Ti GPU utilizing CUDA 10.1 CUDNN 7.6.5 and PyTorch 1.7.0. In each experiment, the initial learning rate is set to 0.01, the final one-cycle learning rate is set to 0.001, the momentum is set to 0.937, the optimizer weight decay is set to 0.0005, and the ship detection confidence is set to 0.7. We use the Stochastic Gradient Descent (SGD) algorithm for learning optimization. The ship instance segmentation task in this research also requires labeling the object instance as supervision information and sending it to the suggested deep learning framework for learning optimization, unlike the general detection task. In order to more thoroughly assess the proposed model, we separated the entire training set into the test set and the training set in a 7:3 ratio. We then compared the detection results with the true value annotation to assess how well the algorithm performed.

4.3 Evaluation metrics

The traditional methods for quantitatively and thoroughly assessing the effectiveness of object detectors are the estimate metrics precision (p), recall (r), intersection of union (IoU), and average precision (AP) (Everingham et al., 2010). The expert

TABLE 1 Information about the SAR imageries in detail for construction.

Dataset	Image (num)	Size (Pixel)	Satellite	Resolution (m)
HRSID (Lin et al., 2017b)	5604	800 x 800	Sentinel-1B/TerraSAR-X /TanDEM-X	1-15
SSDD (Li et al., 2017)	1160	800 x 800	RadarSat-2/TerraSAR-X/Sentinel-1	1-15

The first two SAR image examples in Figure 4 show offshore ships, whereas the last two in the row, respectively, show ships docking in ports and large ships and show the cluster-distributed tiny ships in the canal.

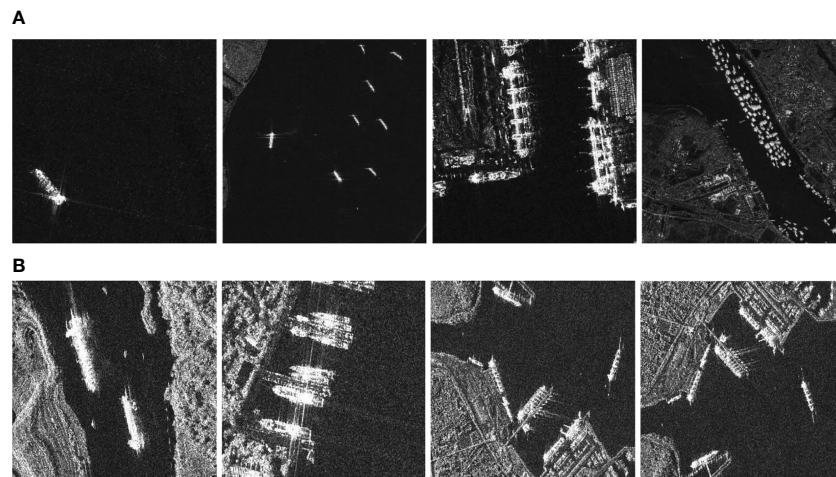


FIGURE 4

Photos are shown from the dataset used in the current paper. (A) some photos from the HRISD dataset and, (B) some photos from SSDD datasets.

annotation of the object's geographic coordinates is referred to as the ground truth in supervised learning for object identification and instance segmentation. The percentage of overlap between the expected outcome and the actual result serves as a proxy for the correlation between two variables; a higher level of overlap denotes a stronger connection and a more precise prediction. Eq (1) states that the bounding box IoU is determined by the percent of overlap between the predicted bounding box and the ground truth bounding box.

The efficiency of various techniques is evaluated using a number of recognized indicators, such as AP, r , p , and IoU, and these indications are particularly specified in the following Eq (1–5) since SAR photo object identification tasks are comparable:

$$IoU_{bbox} = \frac{Bbox_p \cap Bbox_g}{Bbox_p \cup Bbox_g} \quad (1)$$

In object identification tasks, AP is a frequently used indicator that compares the proportion of properly recognized items to the total number of objects in the picture. Another often-used metric is r , which compares the fraction of successfully recognized items to the total number of objects in the picture. It is determined as the ratio of true positives (items that have been accurately identified) to the sum of true positives and false negatives (objects that were present in the image but not detected).

p is an indicator that calculates the proportion of successfully detected items concerning the total number of detected objects in the picture. It is calculated by dividing the number of true positives by the total number of true positives and false positives. IoU (Intersection over Union) is an indicator that calculates the ratio of the intersection of two bounding boxes to the union of two bounding boxes to determine the similarity between two bounding boxes (Bbox p and Bbox g). These indicators (AP, r , p , IoU) are extensively employed in the domain of SAR picture object identification to evaluate and compare the efficacy of various methodologies.

The rate of overlap between the ground mask and predicted mask, as shown in equation (2), determines the mask IoU in a manner similar to how segmentation precision is calculated.

$$IoU_{mask} = \frac{Mask_p \cap Mask_g}{Mask_p \cup Mask_g} \quad (2)$$

The IoU may also be used to assess segmentation tasks such as object recognition in SAR images. The IoU is determined using equation (2), which is comparable to the calculation for IoU of bounding boxes that has been previously described. The IoU mask is the ratio of the predicted mask (Mask p) and the ground truth mask (Mask g) intersection to the union of the two masks. IoU is also known as the Jaccard Index in the context of image segmentation, which is a standard statistic for evaluating the performance of image segmentation algorithms. A high IoU score implies that the predicted mask and the ground truth mask have a high degree of overlap, indicating that the model is accurate.

During classification, algorithms may incorrectly recognize the surroundings and the objects. True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) are the four categorization findings, where TP stands for the number of successfully categorized positive samples, TN for correctly classed negative samples, FN for correctly classified missed positive samples, and FP for correctly classified false alarms in the background. These criteria establish p and r , as shown by equations (3, 4).

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

In classification tasks, the four categorization findings are used to evaluate the algorithm's performance. Precision and recall, two often used indicators in classification tasks, are calculated using TN,

FN, TP, and FP. The equation (3) is used to calculate precision, it calculates the fraction of correctly identified positive samples to the total number of positive samples. A high accuracy score suggests that the algorithm has a low number of false positives, indicating that it accurately identifies a large majority of positive samples.

The AP is established using recall and precision measurements. If the horizontal coordinate is the r value and the vertical coordinate is the precision value, as shown in equation (5), then the area under the recall-precision curve is the AP value in the Cartesian coordinate system:

$$AP = \int_0^1 P(R)dR \quad (5)$$

The mathematical average of all categories in a dataset with multiple classes is defined as the mean AP (mAP). The AP measure is extensively used to assess the effectiveness of object identification systems. The area under the recall-precision curve, which is a plot of recall vs. accuracy, is what it is. According to equation (5), the AP value in the Cartesian coordinate system is the definite integral of the accuracy value with respect to the recall value, ranging from 0 to 1. A greater AP value suggests that the algorithm is doing well, as seen by a larger area under the recall-precision curve.

Mean Average Precision (mAP) is a statistic used to assess the effectiveness of multi-class object identification systems. It is the average of all the AP values in a dataset. It provides an overall measure of the algorithm's performance across all classes in the dataset. A greater mAP number implies that the method performs better across all classes in the dataset.

4.4 Visualization experiment of proposed algorithm

Due to various incident angles of the radar signal, environmental conditions, polarization techniques, etc., the

preprocessing SAR images include clutter noise that interferes with the feature of ships and prohibits ship identification and instance segmentation using CNN. Therefore, while building a SAR dataset for ship identification and instance segmentation, ships should be totally and precisely labeled as opposed to creating an optical RS dataset for object recognition and instance segmentation (Waqas Zamir et al., 2019). In current research work, we have established an effective and reliable algorithm for building an HR-RS dataset for CNN-based ship identification and instance segmentation. Instance segmentation's impacts on low-resolution SAR pictures may be limited in order to escape missing annotation and incorrect annotation brought on by artificial structures that resemble ships (Wang et al., 2019), which are displayed as highlighted spots in low-resolution SAR images. High-resolution remote sensing pictures are utilized to create the dataset, and the images are sliced into 800 x 800 size segments for optimal function development, such as multi-scale training.

The results of ship identification instance segmentation for SAR images using the proposed model are shown in Figures 5, 6. The ground truth mask results are shown in the first row of the figure, and the projected instance outcomes are outcomes presented in the second row. Figures 5, 6 demonstrate how our model's output is suitable for our goal of segmenting instances in HR-RS images. As missed and false alarms increase in our model, instance segmentation is carried out on the mask branch. Finally, these synthetic targets can be detected and segmented quite well, and the segmentation outcomes produced by our model are very close to reality. With the help of our model, the instance segmentation task in HR-RS images was completed successfully.

4.5 Ablation studies

We performed ablation experiments to assess the efficacy of various components in their suggested ship instance segmentation

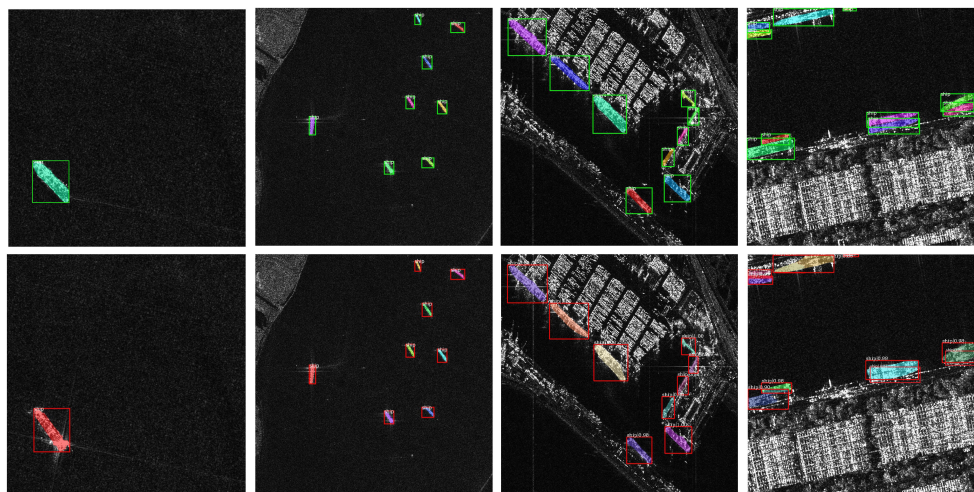


FIGURE 5

Outcomes of the proposed approach instance segmentation in the HRSID dataset (first row show the ground truth and second row is the predicted instance outcomes).

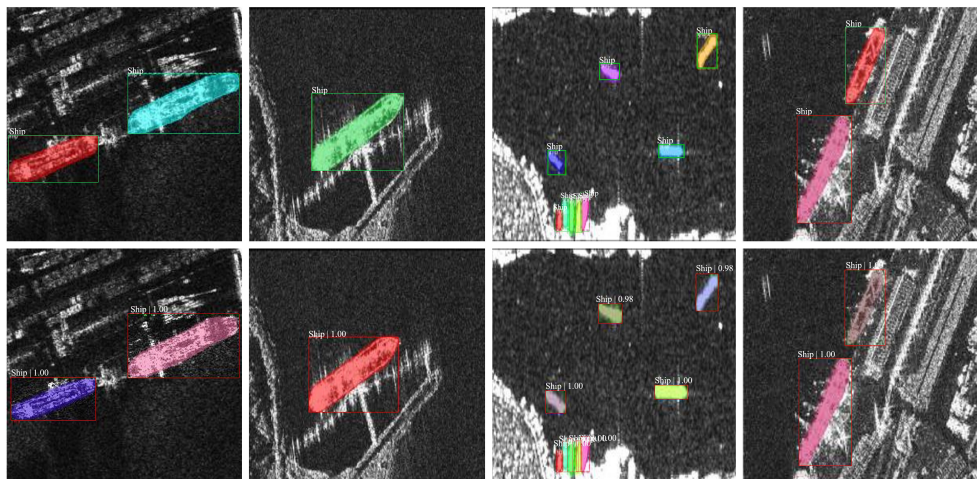


FIGURE 6 Results of the proposed model's instance segmentation in the SSDD detection dataset (the first row show the ground truth and the second row shows the predicted instance outcomes).

detection model. **Table 2** shows the findings of the ablation research. As the default model, the writers used the YOLOv7 model with an input size of 640x640 pixels. The standard model had an AP of 57.8, with an AP50 of 83.7 and an AP75 of 69.5. Also we have added E-ELAN, an edge enhancement module, to the basic model in the first ablation trial. With the inclusion of E-ELAN, the AP increased to 59.4, with an AP50 of 89.6 and an AP75 of 71.9. Then we have added MP-Conv, a multi-path convolution module, to the basic model in the second ablation analysis. The inclusion of MP-Conv increased the AP to 60.7, with an AP50 of 83.9 and an AP75 of 69.8. Cat-Conv, a channel attention transfer convolution module, was added to the baseline model in the third ablation trial. Cat-Conv increased the AP to 62.3, with an AP50 of 83.1 and an AP75 of 68.3. Also we have added SPPSPC, a spatial pyramid pooling module, and convolution to the baseline model in the fourth ablation trial. SPPSPC increased the AP to 63.5, with an AP50 of 87.8 and an AP75 of 73.5. In the last, the authors added all of the previously stated modules (E-ELAN, MP-Conv, Cat-Conv, and SPPSPC) to the baseline model in the fifth and concluding ablation trial. The finished model had the greatest AP of 69.7, as well as an AP50 of 94.9 and an AP75 of 86.5. The authors discovered that incorporating all four modules greatly enhanced the baseline model's performance, particularly in terms of accuracy

and recall, showing the efficacy of their suggested model for real-time ship instance segmentation recognition in complicated backdrop SAR images.

4.6 Comparison with other state-of-the-art techniques

Figures 7 and **8** show the qualitative outcomes of our model and the comparable algorithm on the SSDD and HRSID dataset, individually, to further validate the efficiency of instance segmentation and ship identification. Row 1 displays the ground-truth mask, while rows 2 to 6 display the results of Faster R-CNN, Cascade R-CNN, Mask R-CNN, and Hybrid Task Cascade, respectively. When compared to existing instance segmentation techniques, the results of our improved model can accurately recognize and separate artificial targets in a variety of scenes, as shown in row 7. The expected instance masks, in particular, precisely cover these contrived objectives. As a result of our model's nearly complete elimination of false alarms and missed detections, our mask branch consistently accomplishes superior instance segmentation. When contrast to bounding box identification approaches like Faster R-CNN, Mask R-CNN,

TABLE 2 The ablation experiment study.

Model	Input size	E-ELAN	MP-Conv	Cat-Conv	SPPSPC	AP	AP50	AP75	APS	APM	APL
Yolov7	640x640	-	-	-	-	57.8	83.7	69.5	57.3	60.6	24.5
	640x640	✓	-	-	-	59.4	89.6	71.9	59.1	60.6	39.7
	640x640	-	✓	-	-	60.7	83.9	69.8	56.9	61.2	30.4
	640x640	-	-	✓	-	62.3	83.1	68.3	60.7	63.5	47.8
	640x640	-	-	-	✓	63.5	87.8	73.5	65.5	67.4	45.5
	640x640	✓	✓	✓	✓	69.7	94.9	86.5	73.4	76.8	58.6

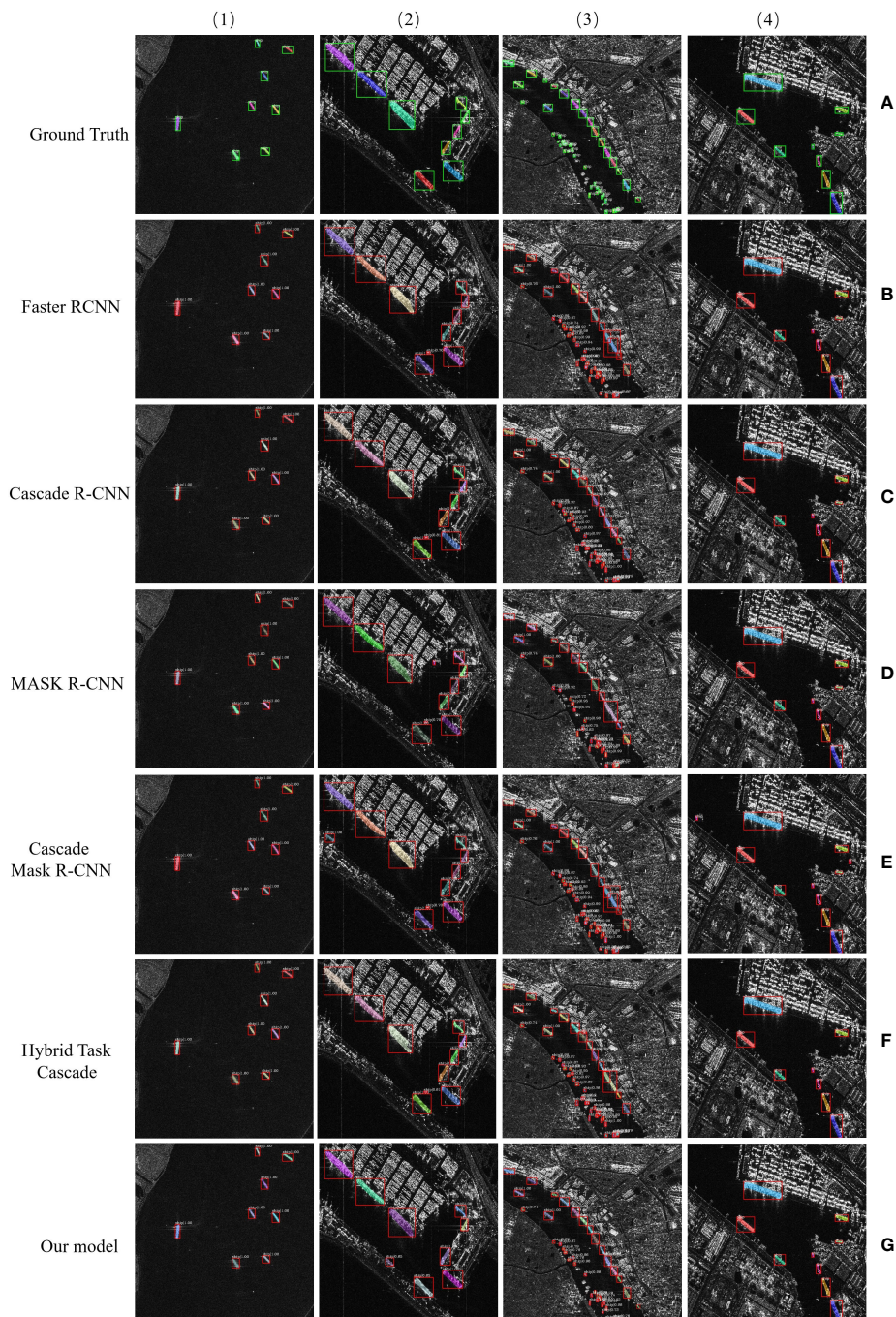


FIGURE 7

Outcomes of CNN-based techniques for visual ship identification instance segmentation using the HRSID dataset. Outcomes from (A) illustrate the ground truth, (B) the Faster-R-CNN technique, (C) the Cascade R-CNN, (D) the Mask R-CNN, (E) the Cascade Mask R-CNN, (F) the Hybrid Task Cascade, and (G) the results from our proposed method.

Cascade Mask R-CNN, Hybrid Task Cascade, and Cascade R-CNN, instance segmentation outcomes are more closely connected to the shape of the original targets. Additionally, separate instances within the same category can be distinguished using the instance segmentation. The ships in Figures 7, 8 stand out because to their dissimilar colors, and in addition, the suggested model, when compared to other instance segmentation approaches, has no false alarms and no missed targets detection while also producing

better results for mask segmentation. The results from the HRSID and SSDD dataset show that our technique is appropriate for instance segmentation in HR-RS photos and outperforms existing instance segmentation strategies when it comes to mask segmentation.

To quantitatively assess the achievement of instance segmentation, we compared the suggested approach with other cutting-edge approaches on the HRSID and SSDD in Tables 3 and

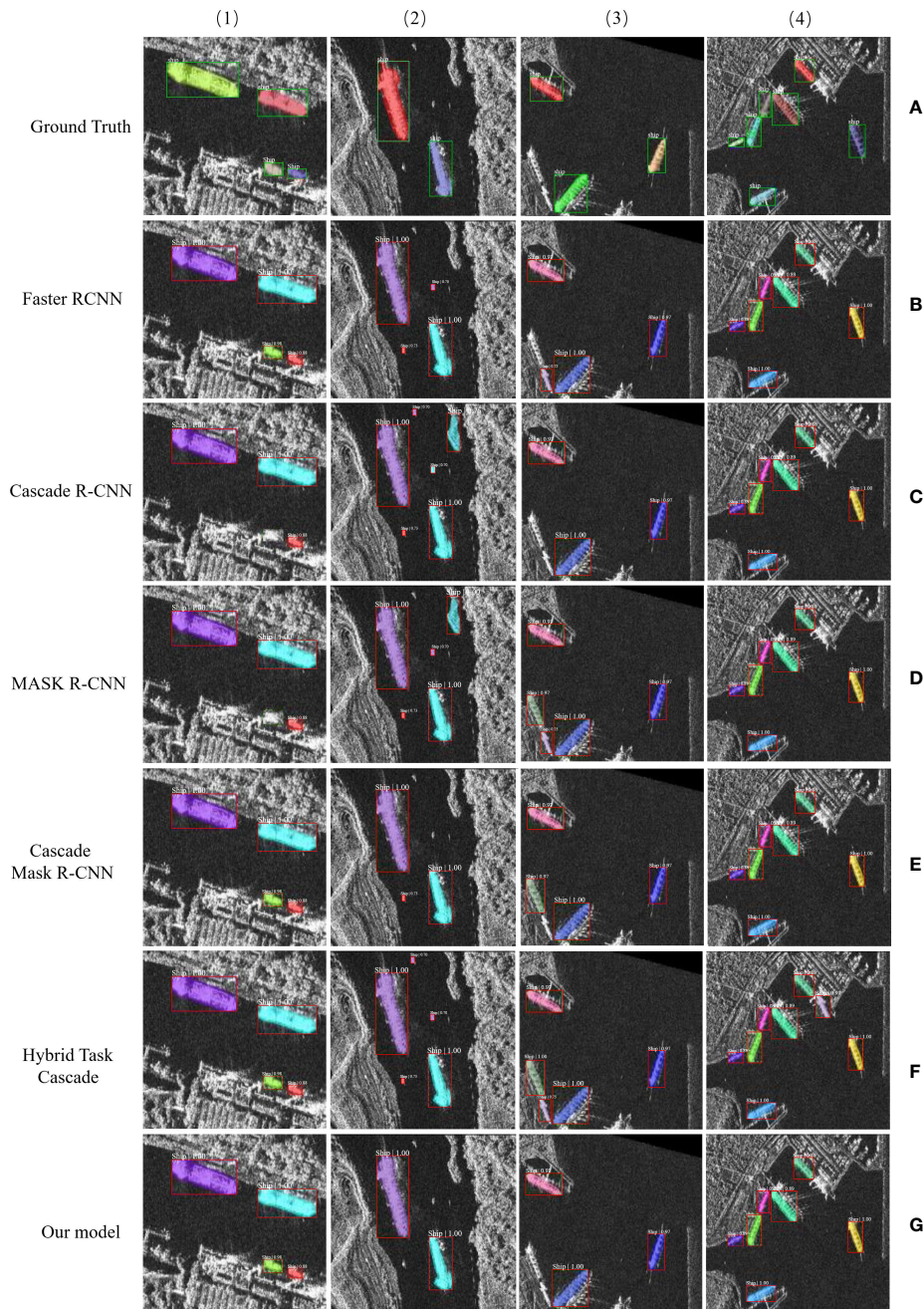


FIGURE 8

Outcomes of CNN-based techniques for visual ship identification instance segmentation using the SSDD. Results from (A) illustrate the ground truth, (B) the Faster-R-CNN technique, (C) the Cascade R-CNN, (D) the Mask R-CNN, (E) the Cascade Mask R-CNN, (F) the Hybrid Task Cascade, and (G) the results from our proposed method.

4. Faster R-CNN, Mask R-CNN, Cascade R-CNN, Cascade Mask R-CNN, and Hybrid Task Cascade are some of these techniques. Tables 3 and 4 show that the suggested strategy achieves the maximum ap of 69.7%. Hybrid Task Cascade and our model outperform Faster R-CNN, Cascade R-CNN, Mask R-CNN, Cascade Mask R-CNN, and Cascade R-CNN by 6.3%, 3.2%, 4.5%, 0.8%, and 2.4%, respectively. In summary, the recommended method has superior instance segmentation effectiveness and better precise predicted instance masks on the HRSID dataset compared to other instance segmentation algorithms. The

reduced parameter count, and computational expense are due to the use of the SiLU activation function, which is more computationally efficient than the traditional ReLU activation function. Additionally, the E-ELAN module selectively weighs the feature maps, further reducing the computational expense without compromising performance. The AP50 score of our model is 94.9%, which is also 10.2% higher than Faster R-CNN, 9.3% higher than Cascade R-CNN, 7.4% higher than Mask R-CNN, 8.2% higher than Cascade Mask R-CNN, and 7.3% higher than Hybrid Task Cascade. Our model achieves an AP75 score of 86.5%, which is an

TABLE 3 Comparing to various cutting-edge methods on the HRSID dataset.

Methods	Backbone	Time (ms)	Model (Size)	AP	AP50	AP75	APS	APM	APL
Faster R-CNN(Ren et al., 2015)	ResNet-50 +FPN	52.6	330M 482M	64.9	84.6	71.5	65.1	66.2	17.8
	ResNet-101 +FPN	64.2		63.4	84.7	71.6	65.7	67.3	25.3
Cascade R-CNN(Cai and Vasconcelos, 2019)	ResNet-50 +FPN	73.9	552M 704M	67.8	85.8	77.6	68.6	68.8	29.9
	ResNet-101 +FPN	85.5		66.5	85.6	77.3	68.2	69.7	28.8
Mask R-CNN (He et al., 2017)	ResNet-50 +FPN	53.7	351M 503M	66.8	87.3	74.9	67.9	67.8	18.4
	ResNet-101 +FPN	62.9		65.2	87.5	74.0	67.3	69.3	24.3
Cascade Mask R-CNN(Cai and Vasconcelos, 2019)	ResNet-50 +FPN	73.0	615M 768M	68.7	86.1	76.6	69.4	68.5	21.5
	ResNet-101 +FPN	87.1		68.9	86.7	76.8	69.8	70.6	22.9
Hybrid Task Cascade (Chen et al., 2019b)	ResNet-50 +FPN	118.9	639M 791M	67.1	88.4	79.8	70.3	72.6	39.0
	ResNet-101 +FPN	134.6		67.3	87.6	79.3	70.8	73.6	32.8
Our Model	ELAN-Net	87	403M	69.7	94.9	86.5	73.4	76.8	58.6

improvement of 14.9% over Faster R-CNN, 9.3% over Cascade R-CNN, 12.5% over Mask R-CNN, 9.7% over Cascade Mask R-CNN, and 7.2% over Hybrid Task Cascade. Mask segmentation has proven to be more precise and superior to other state-of-the-art techniques, such as segmentation utilizing the HRSID dataset. The efficacy of large medium, and small targets on the HRSID dataset has also improved, according to APS, APM, and APL.

Table 3 shows that our model achieves a 70.3% AP, which represents an improvement of 11.7% compared to Faster R-CNN, 9.2% compared to Cascade R-CNN, 13.8% compared to Mask R-CNN, 10.3% compared to Cascade Mask R-CNN, and 2.5% compared to Hybrid Task Cascade. In summary, the recommended model has superior instance segmentation

effectiveness and more precise predicted instance masks when compared to previous instance segmentation algorithms on the SSDD dataset. The AP50 score of our model is also 94.7%, which is an improvement of 15.7% over Faster R-CNN, 3.3% over Cascade R-CNN, 4% over Mask R-CNN, 7.7% over Cascade Mask R-CNN, and 2% over Hybrid Task Cascade. Our model obtains an AP75 of 76.5 percent, which is an improvement of 11% over Faster R-CNN, 9.8% over Cascade R-CNN, 10.8% over Mask R-CNN, 8.8% over Cascade Mask R-CNN, and 1.6% over Hybrid Task Cascade. It has been proven that segmentation using the mask will be more accurate and superior than segmentation using other cutting-edge techniques, such as segmentation on the SSDD dataset. According to APL, APM, and APS, the HRSID dataset’s small, medium, and

TABLE 4 Comparing to various cutting-edge methods on the SSDD dataset.

Methods	Backbone	Time (ms)	Model (Size)	AP	AP50	AP75	APS	APM	APL
Faster R-CNN(Ren et al., 2015)	ResNet-50+FPN ResNet-101+FPN	55.5 66.1	330M 482M	57.5 58.6	78.1 79.0	64.2 65.5	42.8 43.6	57.8 58.1	62.7 61.6
	Cascade R-CNN (Cai and Vasconcelos, 2019)	61.9 70.2		552M 704M	60.7 61.1	90.2 91.4	67.8 66.7	46.4 45.7	61.7 61.4
Mask R-CNN (He et al., 2017)	ResNet-50+FPN ResNet-101+FPN	63.0 72.3	351M 503M	55.3 56.5	91.3 90.7	64.8 65.8	41.8 41.1	55.7 54.4	59.9 60.2
	Cascade Mask R-CNN(Cai and Vasconcelos, 2019)	85.6 93.8		615M 768M	60.2 59.7	88.5 87.2	66.8 67.7	47.5 46.2	63.5 63.0
Hybrid Task Cascade (Chen et al., 2019b)	ResNet-50+FPN ResNet-101+FPN	153.2 168.5	639M 791M	68.7 67.8	91.2 92.6	75.5 74.9	52.2 54.6	68.9 67.8	70.5 73.8
	Our Model	96		403M	70.3	94.7	76.5	55.9	70.2

large target efficacy has also enhanced. We achieve the similar achievement as our model on the NWPU VHR-10 dataset under several AP indicators, and some AP indicators even outperform it.

Tables 3, 4 show how our model performs better with fewer parameters and less computational expense. The proposed model incorporates several improvements to the YOLOv7 backbone architecture, including the addition of an ELAN-Net backbone and FPN, the SiLU activation function, and the E-ELAN module. These improvements allow the model to more effectively extract and use relevant features from SAR images, resulting in improved detection and segmentation performance. Moreover, the proposed model achieves this improved performance while using fewer parameters and less computational expense compared to other modern models, as shown in Tables 3 and 4. The reduced parameter count and computational expense are due to the use of the SiLU activation function, which is more computationally efficient than the traditional ReLU activation function. Additionally, the E-ELAN module selectively weighs the feature maps, further reducing the computational expense without compromising performance.

Furthermore, with comparable model sizes and levels of computational complexity, our models outperform the Mask Scoring R-CNN and Mask R-CNN. Comparing our models to Hybrid Task Cascade and Cascade Mask R-CNN, we find that our models outperform them while consuming less processing power and having a smaller model size. Our network is therefore better than other modern algorithms in terms of model size and processing complexity.

In order to assess the detectors' capacities to locate the ship in complex situations and to test their capacity to deliver adequately observable results, some complex scenarios are added to the datasets. The findings demonstrate that complex situations, like those containing nearby ships and small ships scattered in a cluster, continue to provide a challenge to detectors. The generated mask may accurately show the distribution of ships with their concrete shape pixel-by-pixel with regard to the visual identification outcomes in instance segmentation, laying the groundwork for further instance segmentation investigations. As a result, when compared to other cutting-edge techniques, our model creates instance masks that are more precise and improves the performance of instance segmentation in HR-RS images.

The object detection of RS images has been shown to have problems by CNN. YOLOv7 was actually created as the fundamental detecting network, whereas the ELAN-Net backbone network was designed for advancement. The results of our studies demonstrate that the enhanced algorithm we built would considerably improve the identification efficiency of small-scale items in RS pictures and can increase the accuracy of multi-scale object segmentation. The HRSID and SSDD datasets were used for our investigation because there are no established, open remote sensing mask datasets available, and there might only be a few different varieties. We also need to conduct further research to improve and advance the model inference speed. However, using fuzzy preprocessing techniques to images is also necessary because the processed images are frequently affected by unknown factors (Versaci et al., 2015). Our next study will focus on solving the aforementioned issues, and in order to test our new models, we will first look for and create more RS mask datasets with a wider range

of object classes. Additionally, we will use more accurate and representative datasets. The next phase of our research will involve creating a lightweight framework model that will speed up inference without sacrificing identification accuracy.

In summary, our proposed model achieves better performance with fewer parameters and less computational expense by incorporating several improvements to the YOLOv7 backbone architecture, and by using the SiLU activation function and the E-ELAN module to more effectively extract and use relevant features from SAR pictures.

5 Conclusions

The field of aerospace and remote sensing (RS) domains is heavily influenced by instance segmentation and object recognition tasks, which have a wide range of potential applications in various real-world scenarios. In recent times, the importance of ship identification in RS satellite images has increased. While most current algorithms identify ships using rectangular bounding boxes, they do not segment pixels. As a result, our research offers an enhanced YOLOv7 one-stage detection technique for ship segmentation and identification in RS imagery, capable of accurately recognizing and segmenting ships at the pixel level. We have redesigned the network structure to adapt to the task of ship target segmentation and added two feature optimization modules to the backbone network to increase the robustness of network feature extraction. In addition, we improved the network feature fusion structure and enhanced the prediction capability of multi-scale targets by optimizing the model acceptance domain. Based on the experimental outcomes on the SSDD and HRSID datasets, our model demonstrates improved accuracy in predicting instance masks, promoting the success of instance segmentation in HR-RS imaging and encouraging further advancements in mask prediction accuracy. Our proposed method outperforms existing methods for segmenting ships in remote sensing images, and we plan to extend our research to the segmentation of objects in drone images. While our proposed approach has limitations in handling extremely small or crowded ship instances, we acknowledge this limitation and suggest further optimization of the network architecture and training strategies. Additionally, we have not yet explored the potential of other advanced techniques such as depthwise separable convolution neural network, balance learning, and attention mechanisms, which could be interesting directions for future research. In summary, our proposed approach provides a more precise and effective solution for ship segmentation and identification in RS imagery, and our future work will focus on extending the application of our proposed method to other remote sensing scenarios.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

MY, LZ. Methodology, MY, and SL. software, MY, SL, and LZ. Validation, WJ, SL, and LZ. Formal analysis, MH, QI, and AC. The investigation, ML. Resources, LZ. Data curation, MY, SM, QI, and QY. Writing-original draft preparation, MY. Writing-review and editing, JW, LZ, and SL. Visualization, SL, LZ. Supervision, JW, and SL. Project administration, JW. Funding acquisition, LZ. All authors contributed to the article and approved the submitted version.

Funding

This work is supported by Global atmospheric aerosol dataset development (HX20220168).

References

- Bolya, D., Zhou, C., Xiao, F., and Lee, Y. J. (2019). "Yolact: real-time instance segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*. 9157–9166.
- Cai, Z., and Vasconcelos, N. (2019). Cascade r-CNN: high quality object detection and instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 1483–1498. doi: 10.1109/TPAMI.2019.2956516
- Chang, Y.-L., Anagaw, A., Chang, L., Wang, Y. C., Hsiao, C.-Y., and Lee, W.-H. (2019). Ship detection based on YOLOv2 for SAR imagery. *Remote Sens.* 11, 786. doi: 10.3390/rs11070786
- Chen, S.-W., Cui, X.-C., Wang, X.-S., and Xiao, S.-P. (2021). Speckle-free SAR image ship detection. *IEEE Trans. Image Process.* 30, 5969–5983. doi: 10.1109/TIP.2021.3089936
- Chen, C., He, C., Hu, C., Pei, H., and Jiao, L. (2019a). MSARN: a deep neural network based on an adaptive recalibration mechanism for multiscale and arbitrary-oriented SAR ship detection. *IEEE Access* 7, 159262–159283. doi: 10.1109/ACCESS.2019.2951030
- Chen, S., Mulgrew, B., and Grant, P. M. (1993). A clustering technique for digital communications channel equalization using radial basis function networks. *IEEE Trans. Neural Networks* 4, 570–590. doi: 10.1109/72.238312
- Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., et al. (2019b). "Hybrid task cascade for instance segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4974–4983.
- Chen, H., Sun, K., Tian, Z., Shen, C., Huang, Y., and Yan, Y. (2020). "Blendmask: top-down meets bottom-up for instance segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8573–8581.
- Cui, Z., Li, Q., Cao, Z., and Liu, N. (2019). Dense attention pyramid networks for multi-scale ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* 57, 8983–8997. doi: 10.1109/TGRS.2019.2923988
- Dai, J., He, K., and Sun, J. (2016). "Instance-aware semantic segmentation via multi-task network cascades," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3150–3158.
- Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *Int. J. Comput. Vision* 88, 303–338. doi: 10.1007/s11263-009-0275-4
- Fan, Q., Chen, F., Cheng, M., Lou, S., Xiao, R., Zhang, B., et al. (2019a). Ship detection using a fully convolutional network with compact polarimetric SAR images. *Remote Sens.* 11, 2171. doi: 10.3390/rs11182171
- Fan, W., Zhou, F., Bai, X., Tao, M., and Tian, T. (2019b). Ship detection using deep convolutional neural networks for PolSAR images. *Remote Sens.* 11, 2862. doi: 10.3390/rs11232862
- Gao, F., Shi, W., Wang, J., Yang, E., and Zhou, H. (2019). Enhanced feature extraction for ship detection from multi-resolution and multi-scene synthetic aperture radar (SAR) images. *Remote Sens.* 11, 2694. doi: 10.3390/rs11222694
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*. 2961–2969.
- Hou, B., Ren, Z., Zhao, W., Wu, Q., and Jiao, L. (2019). Object detection in high-resolution panchromatic images using deep models and spatial template matching. *IEEE Trans. Geosci. Remote Sens.* 58, 956–970. doi: 10.1109/TGRS.2019.2942103
- Kang, M., Leng, X., Lin, Z., and Ji, K. (2017). "A modified faster r-CNN based on CFAR algorithm for SAR ship detection," in *2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP): IEEE*. 1–4.
- Kong, W., Liu, S., Xu, M., Yasir, M., Wang, D., and Liu, W. (2023). Lightweight algorithm for multi-scale ship detection based on high-resolution SAR images. *Int. J. Remote Sens.* 44, 1390–1415. doi: 10.1080/01431161.2023.2182652
- Li, J., Qu, C., and Shao, J. (2017). "Ship detection in SAR images based on an improved faster r-CNN," in *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA): IEEE*. 1–6.
- Li, K., Zhang, M., Xu, M., Tang, R., Wang, L., and Wang, H. (2022). Ship detection in SAR images based on feature enhancement swin transformer and adjacent feature fusion. *Remote Sens.* 14, 3186. doi: 10.3390/rs14133186
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017a). "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2117–2125. doi: 10.48550/arXiv.1708.02002
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017b). Focal loss for dense object detection. *Proc. IEEE Int. Conf. Comput. Vision*, 2980–2988. doi: 10.48550/arXiv.1708.02002
- Liu, S., Kong, W., Chen, X., Xu, M., Yasir, M., Zhao, L., et al. (2022). Multi-scale ship detection algorithm based on a lightweight neural network for spaceborne SAR images. *Remote Sens.* 14, 1149. doi: 10.3390/rs14051149
- Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8759–8768.
- Liu, R., Wang, X., Lu, H., Wu, Z., Fan, Q., Li, S., et al. (2021a). SCCGAN: style and characters inpainting based on CGAN. *Mobile Networks Appl.* 26, 3–12. doi: 10.1007/s11036-020-01717-x
- Liu, Y., Zhang, Z., Liu, X., Wang, L., and Xia, X. (2021b). Efficient image segmentation based on deep learning for mineral image classification. *Advanced Powder Technol.* 32, 3885–3903. doi: 10.1016/j.apt.2021.08.038
- Mou, L., and Zhu, X. X. (2018). Vehicle instance segmentation from aerial image and video using a multitask learning residual fully convolutional network. *IEEE Trans. Geosci. Remote Sens.* 56, 6699–6711. doi: 10.1109/TGRS.2018.2841808
- Nie, X., Duan, M., Ding, H., Hu, B., and Wong, E. K. (2020). Attention mask r-CNN for ship detection and segmentation from remote sensing images. *IEEE Access* 8, 9325–9334. doi: 10.1109/ACCESS.2020.2964540
- Qian, X., Lin, S., Cheng, G., Yao, X., Ren, H., and Wang, W. (2020). Object detection in remote sensing images based on improved bounding box regression and multi-level features fusion. *Remote Sens.* 12, 143. doi: 10.3390/rs12010143
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 28. doi: 10.48550/arXiv.1506.01497
- Shao, Z., Zhang, X., Zhang, T., Xu, X., and Zeng, T. (2022). RBFA-net: a rotated balanced feature-aligned network for rotated SAR ship detection and classification. *Remote Sens.* 14, 3345. doi: 10.3390/rs14143345

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Su, N., He, J., Yan, Y., Zhao, C., and Xing, X. (2022). SII-net: spatial information integration network for small target detection in SAR images. *Remote Sens.* 14, 442. doi: 10.3390/rs14030442
- Su, H., Wei, S., Yan, M., Wang, C., Shi, J., and Zhang, X. (2019). "Object detection and instance segmentation in remote sensing imagery based on precise mask r-CNN," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium: IEEE*. 1454–1457.
- Sun, Z., Dai, M., Leng, X., Lei, Y., Xiong, B., Ji, K., et al. (2021a). An anchor-free detection method in high-resolution SAR images based on YOLOv5," in *2022 Photonics & Electromagnetics Research Symposium (PIERS): IEEE*. 647–653.
- Sun, Z., Lei, Y., Leng, X., Xiong, B., and Ji, K. (2022). "An improved oriented ship detection method in high-resolution SAR image based on YOLOv5," in *2022 Photonics & Electromagnetics Research Symposium (PIERS): IEEE*. 647–653.
- Sun, Z., Leng, X., Lei, Y., Xiong, B., Ji, K., and Kuang, G. (2021b). BiFA-YOLO: a novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images. *Remote Sens.* 13, 4209. doi: 10.3390/rs13214209
- Versaci, M., Calcagno, S., and Morabito, F. C. (2015). "Fuzzy geometrical approach based on unit hyper-cubes for image contrast enhancement," in *2015 IEEE international conference on signal and image processing applications (ICSIPA): IEEE*. 488–493.
- Wang, X., Kong, T., Shen, C., Jiang, Y., and Li, L. (2020a). "Solo: segmenting objects by locations," in *European Conference on Computer Vision*. 649–665.
- Wang, C.-Y., Liao, H.-Y. M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., and Yeh, I.-H. (2020b). "CSPNet: a new backbone that can enhance learning capability of CNN," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 390–391.
- Wang, J., Lu, C., and Jiang, W. (2018). Simultaneous ship detection and orientation estimation in SAR images based on attention module and angle regression. *Sensors* 18, 2851. doi: 10.3390/s18092851
- Wang, Y., Wang, C., Zhang, H., Dong, Y., and Wei, S. (2019). Automatic ship detection based on RetinaNet using multi-resolution gaofen-3 imagery. *Remote Sens.* 11, 531. doi: 10.3390/rs11050531
- Waqas Zamir, S., Arora, A., Gupta, A., Khan, S., Sun, G., Shahbaz Khan, F., et al. (2019). "Isaid: a large-scale dataset for instance segmentation in aerial images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 28–37.
- Wei, S., Su, H., Ming, J., Wang, C., Yan, M., Kumar, D., et al. (2020). Precise and robust ship detection for high-resolution SAR imagery based on HR-SDNet. *Remote Sens.* 12, 167. doi: 10.3390/rs12010167
- Wu, Y., Sheng, H., Zhang, Y., Wang, S., Xiong, Z., and Ke, W. (2022). Hybrid motion model for multiple object tracking in mobile devices. *IEEE Internet Things J.* doi: 10.1109/JIOT.2022.3219627
- Xie, E., Sun, P., Song, X., Wang, W., Liu, X., Liang, D., et al. (2020). "Polarmask: single shot instance segmentation with polar representation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 12193–12202.
- Xiong, B., Sun, Z., Wang, J., Leng, X., and Ji, K. (2022). A lightweight model for ship detection and recognition in complex-scene SAR images. *Remote Sens.* 14, 6053. doi: 10.3390/rs14236053
- Xu, X., Feng, Z., Cao, C., Li, M., Wu, J., Wu, Z., et al. (2021). An improved swin transformer-based model for remote sensing object detection and instance segmentation. *Remote Sens.* 13, 4779. doi: 10.3390/rs13234779
- Xu, X., Zhang, X., Shao, Z., Shi, J., Wei, S., Zhang, T., et al. (2022a). A group-wise feature enhancement-and-Fusion network with dual-polarization feature enrichment for SAR ship detection. *Remote Sens.* 14, 5276. doi: 10.3390/rs14205276
- Xu, X., Zhang, X., Zhang, T., Yang, Z., Shi, J., and Zhan, X. (2022b). Shadow-Background-Noise 3D spatial decomposition using sparse low-rank Gaussian properties for video-SAR moving target shadow enhancement. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2022.3223514
- Yasir, M., Jianhua, W., Mingming, X., Hui, S., Zhe, Z., Shanwei, L., et al. (2022). Ship detection based on deep learning using SAR imagery: a systematic literature review. *Soft Computing* 1–22. doi: 10.1007/s00500-022-07522-w
- Yasir, M., Jianhua, W., Mingming, X., Hui, S., Zhe, Z., Shanwei, L., et al. (2023a). Ship detection based on deep learning using SAR imagery: a systematic literature review. *Soft Computing* 27, 63–84. doi: 10.1007/s00500-022-07522-w
- Yasir, M., Shanwei, L., Xu, M., Sheng, H., Hossain, M. S., Colak, A. T. I., et al. (2023b). Multi-scale ship target detection using SAR images based on improved Yolov5. *Front. Mar. Science*. doi: 10.3389/fmars.2022.1086140
- Yin, M., Zhu, Y., Yin, G., Fu, G., and Xie, L. (2022). Deep feature interaction network for point cloud registration, with applications to optical measurement of blade profiles. *IEEE Trans. Ind. Informatics*. doi: 10.1109/TII.2022.3220889
- Zeng, X., Wei, S., Shi, J., and Zhang, X. (2021). A lightweight adaptive roi extraction network for precise aerial image instance segmentation. *IEEE Trans. Instrumentation Measurement* 70, 1–17. doi: 10.1109/TIM.2021.3121485
- Zhang, Y., Guo, L., Wang, Z., Yu, Y., Liu, X., and Xu, F. (2020c). Intelligent ship detection in remote sensing images based on multi-layer convolutional feature fusion. *Remote Sens.* 12, 3316. doi: 10.1016/j.isprsjprs.2020.05.016
- Zhang, J., Lin, S., Ding, L., and Bruzzone, L. (2020a). Multi-scale context aggregation for semantic segmentation of remote sensing images. *Remote Sens.* 12, 701. doi: 10.3390/rs12040701
- Zhang, H., Luo, G., Li, J., and Wang, F.-Y. (2021a). C2FDA: coarse-to-fine domain adaptation for traffic object detection. *IEEE Trans. Intelligent Transportation Syst.* 23, 12633–12647. doi: 10.1109/JSTARS.2021.3102989
- Zhang, S., Wu, R., Xu, K., Wang, J., and Sun, W. (2019a). R-CNN-based ship detection from high resolution remote sensing imagery. *Remote Sens.* 11, 631. doi: 10.3390/rs11060631
- Zhang, T., Zeng, T., and Zhang, X. (2023). Synthetic aperture radar (SAR) meets deep learning. *Remote Sens.* 15, 2.
- Zhang, T., and Zhang, X. (2021b). "Integrate traditional hand-crafted features into modern CNN-based models to further improve SAR ship classification accuracy," in *2021 7th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR): IEEE*. 1–6.
- Zhang, T., and Zhang, X. (2019). High-speed ship detection in SAR images based on a grid convolutional neural network. *Remote Sens.* 11, 1206. doi: 10.3390/rs11101206
- Zhang, T., and Zhang, X. (2021a). Injection of traditional hand-crafted features into modern CNN-based models for SAR ship classification: what, why, where, and how. *Remote Sens.* 13, 2091. doi: 10.3390/rs13112091
- Zhang, T., and Zhang, X. (2022a). A full-level context squeeze-and-excitation ROI extractor for SAR ship instance segmentation. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2022.3166387
- Zhang, T., and Zhang, X. (2022b). HTC+ for SAR ship instance segmentation. *Remote Sens.* 14, 2395. doi: 10.3390/rs14102395
- Zhang, T., and Zhang, X. (2022c). A mask attention interaction and scale enhancement network for SAR ship instance segmentation. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2022.3189961
- Zhang, T., and Zhang, X. (2022d). A polarization fusion network with geometric feature embedding for SAR ship classification. *Pattern Recognition* 123, 108365. doi: 10.1016/j.patcog.2021.108365
- Zhang, T., Zhang, X., Liu, C., Shi, J., Wei, S., Ahmad, I., et al. (2021b). Balance learning for ship detection from synthetic aperture radar remote sensing imagery. *ISPRS J. Photogrammetry Remote Sens.* 182, 190–207.
- Zhang, T., Zhang, X., Shi, J., and Wei, S. (2019b). Depthwise separable convolution neural network for high-speed SAR ship detection. *Remote Sens.* 11, 2483. doi: 10.3390/rs11212483
- Zhang, T., Zhang, X., Shi, J., and Wei, S. (2020b). HyperLi-net: a hyper-light deep learning network for high-accurate and high-speed ship detection from synthetic aperture radar imagery. *ISPRS J. Photogrammetry Remote Sens.* 167, 123–153. doi: 10.1109/TGRS.2022.3167569
- Zhou, W., Liu, J., Lei, J., Yu, L., and Hwang, J.-N. (2021). GMNet: graded-feature multilabel-learning network for RGB-thermal urban scene semantic segmentation. *IEEE Trans. Image Process.* 30, 7790–7802. doi: 10.1109/TIP.2021.3109518
- Zhou, W., Lv, Y., Lei, J., and Yu, L. (2019). Global and local-contrast guides content-aware fusion for RGB-d saliency prediction. *IEEE Trans. Systems Man Cybernetics: Syst.* 51, 3641–3649. doi: 10.1109/TSMC.2019.2957386
- Zhou, W., Wang, H., and Wan, Z. (2022b). Ore image classification based on improved CNN. *Comput. Electrical Eng.* 99, 107819. doi: 10.1016/j.compeleceng.2022.107819
- Zhou, G., Yang, F., and Xiao, J. (2022a). Study on pixel entanglement theory for imagery classification. *IEEE Trans. Geosci. Remote Sens.* 60, 1–18.
- Zhu, H., and Zhao, R. (2022). Isolated Ni atoms induced edge stabilities and equilibrium shapes of CVD-prepared hexagonal boron nitride on the Ni (111) surface. *New J. Chem.* 46, 17496–17504. doi: 10.1039/D2NJ03735A
- Zong, C., and Wan, Z. (2022). Container ship cell guide accuracy check technology based on improved 3D point cloud instance segmentation. *Brodogradnja: Teorija i praksa brodogradnje i pomorske tehnike* 73, 23–35. doi: 10.21278/brod73102
- Zong, C., and Wang, H. (2022). An improved 3D point cloud instance segmentation method for overhead catenary height detection. *Comput. electrical Eng.* 98, 107685. doi: 10.1016/j.compeleceng.2022.107685
- Zou, L., Zhang, H., Wang, C., Wu, F., and Gu, F. (2020). Mw-acgan: generating multiscale high-resolution SAR images for ship detection. *Sensors* 20, 6673. doi: 10.3390/s20226673