# A System to accurately detect and categorize Maritime Targets in Synthetic Aperture Radar Images.pdf

*by* VBIT LIB

# A System to accurately detect and categorize Maritime Targets in Synthetic Aperture Radar Images

Dr. K. Shirisha Reddy
Associate Professor
Vignana Bharathi Institute of Technology
kshirishacse@vbithyd.ac.in

Pinjala Siva Teja
Undergraduate
Vignana Bharathi Institute of Technology
pinjalasivateja40@gmail.com

Venkata Ramana Panigrahi
Undergraduate
Vignana Bharathi Institute of Technology
venkataramanapanigrahi1015@gmail.com

***Abstract*: Synthetic Aperture Radar technology uses an active remote sensing technique to capture images by taking advantage of long-range propagation characteristics of radar signals. This technology can capture satellite imagery regardless of weather and lighting conditions. Object and Target detection is vital for reconnaissance activities and is a major application of remote sensing technology in the military fields. Maritime targets in Synthetic Aperture Radar images contain characteristically uncertain contour information, a complex background, and display high speckles. Deep-learning models have evolved for object detection and other fields of computer vision. The SAR Maritime target detection and classification is done through many object detection algorithms and Convolution Neural Networks achieved good results but they had many false detections, many missing features and inaccurate classification. To address this issue and for more accuracy, we attempted to implement the Single Shot MultiBox Detector (SSD), a deep learning model that is widely used for object detection, and attempted to implement it with a Residual Network for better feature fusion. This proposed method. augmented the performance of maritime target detection in SAR images in enhancing the detection accuracy and speed.**

**Keywords: Synthetic Aperture Radar (SAR), Remote Sensing, Object detection, Maritime targets, Deep Learning, SSDD, Detection, Classification.**

## I. Introduction

Synthetic Aperture Radar (SAR) is a technology that uses remote sensing techniques with usage of microwave signals to image Earth. Remote Sensing involves acquiring data on objects or events from afar, eliminating the need for direct contact, unlike direct observation methods. Synthetic Aperture Radar (SAR) is a type of radar which creates two-dimensional images or three-dimensional restorations of objects like landscapes. An active system that sends a microwave signal from a sensor platform to the ground and picks up backscattered waves that are reflected back to a receiver built into the payload is responsible for creating the images. SAR images make use of microwave frequencies; thus, SAR systems gives unique images exhibiting the electrical and geometrical properties of a surface under all climatic environments. SAR images are capable of providing their own illumination; thus, they can be in day or night-light conditions. [1]

SAR images are used in various applications, such as topographic mapping and classification of land, retrieval of various parameters, object detection and recognition, the SAR images are widely used for object and target detection in military and civilian fields. SAR images are highly considered in research areas to locate and identify objects in images, including visual and Synthetic Aperture Radars. [2] The SAR Images are important for target recognition and classification, and contribute to overall safety. Maritime target detection is an important technology using SAR Images, supporting water traffic monitoring and marine safety maintenance.[3]

SAR Maritime target detection is already being performed using many techniques, but they are not capable of extracting features from the data. Maritime target detection is a crucial task for marine environment monitoring for finding of illegal fishing activities, oil spills and management of maritime traffic. The existing SAR Maritime target detection techniques make use of traditional approaches and Convolutional Neural Networks (CNN). Traditional approaches use a constant false-alarm rate (CFAR). They try to compute the threshold value, and by comparing the value of threshold and the false alarm rate, they detect maritime target pixels, and the remaining pixels are considered as the background. This method fails for offshore images, where there are many false positives, leading to inaccurate detections. The Convolution Neural Networks are widely used to recognize patterns that are further used for object detection, class identification, and image analysis. In recent years, several methods based on Deep Neural Networks (DNNs) have been implemented to improve the performance of SAR maritime target detection. [4]

Maritime target detection in Synthetic Aperture Radar images is now highly prioritized. Deep Learning has led to notable performance in various object detection tasks. [5] This system introduces a detection model based on a single-shot multibox detector using Residual Networks. The model was built to train it on recognizing and localizing objects in an image. The SSD is purely a convolutional neural network that can be organized into three parts: A Base Convolution Network, Auxiliary Convolution Network and Prediction Convolution Network. Here, we use residual networks as the base convolution neural network for extracting the feature map and to provide lower-level feature maps. An Auxiliary Convolution Network was implemented along with the base network to produce higher-level proposals from the feature map. The proposals can be considered positive outcomes. These positive outcomes are further considered by the Prediction Convolution Network for categorizing maritime targets and plotting bound boxes. A single-shot multibox detector is a unified framework for object and target detection with various layered networks.

The use of Residual Networks or ResNets as the backbone addresses the vanishing gradient problem and is suitable for feature fusion, allowing the model to learn more complex features from SAR images. Transfer Learning can also be used for this proposed model, as SSD models are pretrained on natural images and can be fine-tuned for learning and extracting the features from SAR data. [6]

## II. Literature Survey

Lilian Asimwe Leonidas and Yang Jie [7] introduced an improved CNN architecture for classifying ships into five categories within an Intelligent Transport System. The model was fine-tuned using pre-trained models, such as AlexNet and ResNet. SSENet [8] focused on extracting the ship size from SAR images, which is a crucial parameter for classification. A novel CNN architecture called SSENet was proposed to accurately estimate ship dimensions. The outputs are the model's estimated scaled length and width. The scaled values were rescaled to regular levels. The results indicate that the proposed deep learning approach considerably improves the size estimation accuracy, thereby increasing the reliability of maritime traffic monitoring and analysis. Yan and Wang [9] presented an ensemble learning approach for ship classification, combining multiple classifiers and incorporating Automatic Identification System (AIS) data for improved accuracy. DCPN [10] proposed a Dual-Stream Contrastive Predictive Network (DCPN) that fuses deep and handcrafted features for ship classification. The network was constructed with two streams: feature streaming and contrastive prediction streaming. Li and Chen [11] provided a comprehensive survey of real-time ship detection methods using deep learning in SAR images. This study compared different techniques, pretrained models and datasets. Using real-time object detection techniques and pre-trained models for real-time SAR ship recognition, the primary goal of this survey was to create a deep learning-based real-time SAR ship detection model. SAR-SHIPNET [12] introduced SAR-ShipNet, a CNN architecture that incorporates bidirectional coordinate attention and multiresolution feature fusion for ship detection in complex maritime scenarios. Experiments have shown that SAR-ShipNet results in competitive detection performance, outperforms traditional detection methods, and provides a more accurate and reliable maritime surveillance system. AMANet [13] introduced an adaptive multi hierarchical attention network designed for detecting small and coastal ships in complex SAR environments. The AMANet was developed by integrating the AMAM between the backbone network and the FPN, showing its ability to be easily incorporated into various frameworks to enhance object detection performance. Yao Chen, Tao Duan, Changyuan Wang, Yuanyuan Zhang and Mo Huang [14] present a multiscale ship detection method using deep CNNs, emphasizing speed and accuracy through a concise and efficient architecture. The feature extraction module applies DarkNet-53 as the backbone, adopted to fuse multiscale features. It also fuses several strategies, particularly soft-NMS, augmentation of mosaic data, training on multiscale and hybrid optimization.

## III. Design and Dataset

### A. Dataset

This experiment utilized the SAR Ship Detection Dataset (SSDD) [14]. It is one of the most popular datasets used on ship detection with deep learning methods and Synthetic Aperture Radar images. The set consisted of 1,160 images with three different annotations: polygon segmentation, rotatable bounding boxes and, bounding boxes. The data set comes with JPG as well as XML files for the annotation. In the SSDD dataset, there are SAR picture samples with various features. These include many polarization kinds (HH, VV, VH, and HV), photos taken by various sensors like RadarSat-2, TerraSAR-X, and Sentinel-1, and resolutions ranging from 1 to 15 meters. A range of sea conditions, ship situations (both in onshore and offshore), and vessel sizes are also included in the dataset.[15]

### B. Model Architecture

The code implements a single-shot multibox detector (SSD) architecture using ResNet as the base network. SSD is a popular architecture for object detection because it performs detection in a single pass-through network.

This is a brief overview of the model.

The backbone network uses a pretrained Resnet 101 network as a feature extractor. ResNet is chosen for its ability to effectively capture hierarchical features of varying scales due to its residual block architecture. The architecture of ResNet incorporates multiple layers arranged in blocks. Each block is comprised of numerous convolutional layers, which are accompanied by activation functions. [16]

In the second stage, a block of auxiliary convolutional layers is utilized, to extract the features at several scales and gradually decreasing the size of the input to each consequent layer.

The prediction layers in the SSD determine most of the single-shot strategies for object detection. The SSD directly predicts the class scores and offsets of bounding boxes at different scales and aspect ratios, efficiently computing the detection outputs in real time. In this way, prediction layers enable SSD to deal effectively with the high diversity in object sizes and configurations. [17]

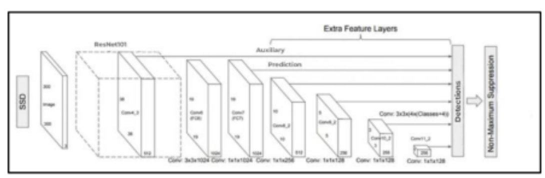


Fig 3.1 Model Architecture.

### C. Proposed Backbone Structure

The SSD model directly predicts the object locations and class scores from the base network one-shot forwarding by designing mechanisms to achieve a good speed and balance index between the model accuracy and available computational resources.

Feature Extractor with ResNet-101

The first step in the SSD model was to define a feature extractor using the ResNet-101 architecture. Hence, ResNet-101 can be considered a pre-trained deep convolutional neural network over a very large dataset to learn very rich feature representations. In the current implementation, the early layers of ResNet-101 are reused to process the input images to generate feature maps, which then serve as the base for object detection at different scales. This consists of a class of feature extractors from the initial layers of ResNet-101, which are used to create a backbone network. This ensured that the extracted features were highly informative. The base network is used to obtain the low-level feature maps used by the Auxiliary Network. [18]

## IV. Implementation

The dataset was cleaned and labelled again into three categories: boat, ship and frigate, by detecting the size and features of the maritime targets. The dataset contains 304 boats, 584 ship and 291 frigate targets. The annotations of the images were as follows: width, height, and depth. Name, pose, truncated, difficult and bounding box values of the object in the image. With the usage of "labelimg" [19] the objects in the image files were labelled in Pascal VOC format which generated the xml label files The bounding box format was adjusted to XYXY format. Images with high speckle were removed, and a dataset of 350 images was created. The dataset was further categorized as the Train, Val and Test datasets.

### A. Proposed System

The system is designed in such a way that it takes the user input of SAR images acquired from the satellites resources and processes it through the built SSD model, detects the bounding boxes around the predicted maritime target, and categorizes them.

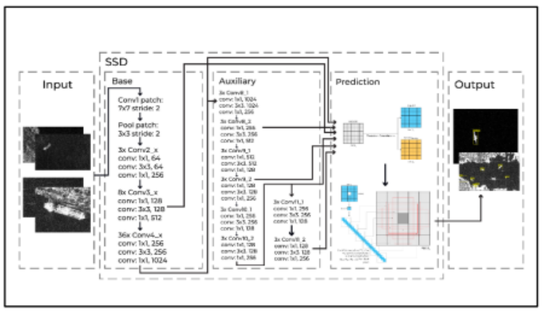


Fig. 4.1. Proposed System Architecture.

### B. Maritime Target Detection Model

The implementation of an overall framework for training an object detection model in single-shot multi-box detector architecture, particularly SSD300, using PyTorch. It involves data preparation, model usage and training procedures to effectively learn from and infer the annotated image datasets. [20]

The script imports all libraries and modules that will be used during training. It configures settings in respect of data, model architecture and training with respect to paths of data and checkpoints, learning rates, batch sizes, and other hyper-parameters. It configures the data folder, batch size during training, number of iterations and parameters like learning rate and momentum. A model and an optimizer are initialized in the 'main' function.

In case a checkpoint is given, it will load the status of the model and the optimizer state from the checkpoint. This ensures that training picks up exactly where one left off. Otherwise, if no checkpoint is given, a new SSD300 model will be created with an optimizer configured with different learning rates for biases and non-bias parameters. We will first prepare the dataset to be used in training with the PascalVOCDataset class, likely to load and preprocess object detection data. Then, it creates a DataLoader for batching and shuffling of the data. Configures a DataLoader to deal with a number of workers for the loading of the data efficiently and includes a custom collate function for the handling of batches having variable-sized data. The training loop may be decomposed into epochs. The learning rate can be changed after every epoch, by giving the user-defined decay points. The 'train' function manages training at each epoch by updating the model weights after forward and backward passes through a network. It handles the calculation of loss computation, clipping gradients if specified and the update of model parameters. It also reports training statistics, like batch time, data loading time, and loss, at periodic intervals. The script saves all model and optimizer state in a checkpoint file after every epoch. This comes very handy in case you want to resume training later or to evaluate model performance. Utility functions are used to outline memory and performance, it also records time taken for data loading and model training; free memory by deleting variables no longer needed.

The first step in the SSD model was to define a feature extractor using the ResNet-101 architecture. Hence, ResNet-101 can be considered a pre-trained deep convolutional neural network over a very large dataset to learn very rich feature representations. In the current implementation, the early layers of ResNet-101 are reused to process the input images to generate feature maps, which then serve as the base for object detection at different scales. This consists of a class of feature extractors from the initial layers of ResNet-101, which are used to create a backbone network.

This ensured that the extracted features were highly informative. The base network is used to get the low-level feature maps which are used by the Auxiliary Network.

Here, additional convolutional layers are added, referred to as the auxiliary convolutions. This will further make the model more efficient with respect to the detection of scaled objects. It can make feature maps at different levels of scale, so both small and large-level object detection are possible. Auxiliary convolutions are dependent on the feature maps out coming from the ResNet-101 backbone and further processed to obtain the final set of feature maps used for detection.

Prediction layers: final layers are responsible for predicting object bounding boxes and their classification. Class scores and box coordinates are outlining the probability of every class in a default box and the outline of detected objects. Prediction layers work upon the feature maps produced by the backbone and auxiliary convolutions, enabling the model to make predictions at every scale.

Two significant components constitute the loss function during the training of the SSD model: the localization loss and the confidence loss. The localization loss is a computation of how ground truth compares with the smooth L1 loss functions in the predicted bounding box related to some ground truth. The other measure is the confidence loss, done by cross-entropy loss based on the predicted class scores compared to actual class labels. These losses are summed into the global loss and minimized in the process of training to optimize the model.

### C. *Training and Fine Tuning*

The model is trained for 20,000 epochs with 496 images and batch size 16. The SAR images were of different classes and resolutions. When training a neural network, several hyper parameters play a crucial role. The worker's parameter controls the number of data loading processes, The learning rate influences the speed of weight updates. Additionally, final learning rate value after decay is also fine-tuned. The momentum term accelerates gradient descent, and weight_decay helps prevent overfitting. At test time, the model will simply take in the input images, perform the class scoring with respective bounding box prediction at each location, and each default box will then be filtered out according to the confidence threshold in an attempt to discard those low-confidence detections. Non-maximum suppression discards all redundant detections by simply keeping a maximal set of those confident predictions. The output from here comes in the form of detected objects' class labels and bounding box coordinates ready for evaluation or further processing.

## V. RESULTS AND VALIDATION

The system was tested for few onshore and offshore satellite images and was able to accurately detect and categorize the maritime targets and was capable to detect them as 'boat', 'ship' and 'frigate' with utmost accuracy and efficiency. The system demonstrated exceptional accuracy and competence in detecting and categorizing maritime targets as boats, ships, or frigates within Synthetic Aperture Radar imagery. A dataset which consists of 50 images with 150 ship targets was created for evaluation.

The system performance was evaluated and compared after 20,000 epochs and 50,000 epochs based on a Precision (mAP), Recall and f1-Score. The following are results after the system was evaluated on testing data with 50 images with 150 detection targets from the SSDD dataset. While the system excels in detecting maritime targets of a smaller range, further improvements can be explored to enhance performance on small and occluded maritime targets. The evaluation metrics can be improved by training model for a greater number of epochs.

| **Training Epochs** | **Evaluation Metrics** | | |
|---|---|---|---|
| | ***Precision (mAP)*** | ***Recall*** | ***F1-Score*** |
| 20,000 | 75 | 78 | 77 |
| 50,000 | 86 | 87 | 85 |

TABLE I. EVALUTION METRICS

## VI. CONCLUSION AND FUTURE ENHANCEMENT

In conclusion, SSD with a ResNet 101 backbone network augments the strengths of both SSD's single-shot detection approach and ResNet's powerful feature extraction capabilities. The ResNet backbone provides robust feature maps at multiple scales, which are crucial for detecting objects of different sizes and shapes. By directly predicting bounding boxes and class probabilities from these feature maps, SSD achieves real-time object detection performance suitable for various applications.

The overall architecture of the SSD model with ResNet-101 is a well-organized and effective pipeline for object detection, where the strong feature extractor is pipelined with further added convolutional layers to generate multi-scale feature maps and thus, be able to accurately detect objects at varied scales. Training and inference are both supported in this model to allow the correct detection of objects within context. The model is well implemented using PyTorch framework and Single Shot MultiBox Detector model architecture.

The future enhancement can be optimizing the dataset by evenly distributing the classes evenly of ship targets and training it for more epochs for better results. We can incorporate different categories of ships and train it on a dataset consisting of various classes of ship targets. Potential avenues for future research include incorporating attention mechanisms and also by exploring alternative backbone architectures.

We can even implement the system to work on real SAR images which aren't preprocessed to test the accuracy of the model. In the used dataset there are images of different bands thus we can try to implement the system based on the band for more effective results.

## REFERENCES


[1] "What is Synthetic Aperture Radar" by Earthdata Learn Backgrounders, NASA.

[2] Reigber A, Scheiber R, Jager M, et al. "Very-high-resolution airborne synthetic aperture radar imaging: Signal processing and applications" [J]. Proceedings of the IEEE, 2013, 101(3): 759-783.

[3] "SAR 101: An Introduction to Synthetic Aperture Radar," by Daniel Hogan.

[4] Y. Wang, C. Wang, and H. Zhang, "Ship classification in high-resolution SAR images using deep learning of small datasets".

[5] "End-to-End Ship Detection in SAR Images for Complex Scenes Based on Deep CNNs" by Yao Chen, Tao Duan, Changyuan Wang, Yuanyuan Zhang and Mo Huang.

[6] Eddie Forson "Understanding SSD MultiBox - Real-Time Object Detection in Deep Learning" published in Towards Data Science.

[7] "Ship Classification Based on Improved Convolutional Neural Network Architecture for Intelligent Transport System" by Lilian Asimwe Leonidas and Yang Jie.

[8] Yibin Ren, Xiaofeng Li and Huan Xu, "Extracting Ship's Size from SAR Images by Deep Learning," ResearchGate, Dec. 2022.

[9] Z. Yan, X. Song, L. Yang, and Y. Wang, "Ship classification in synthetic aperture radar images based on multiple classifiers ensemble learning and automatic identification system data transfer learning".

[10] X. Feng, H. Zheng, Z. Hu, L. Yang, and M. Zheng, "Dual-stream contrastive predictive network with joint handcrafted feature view for SAR ship classification".

[11] Jianwei Li and Jie Chen, "A Survey on Deep-Learning-Based Real-Time SAR Ship Detection," Remote Sens., vol. 11, no. 7, p. 765, Apr. 2019.

[12] Y. Deng, D. Guan, Y. Chen, W. Yuan, J. Ji, and M. Wei, "SAR-SHIPNET: SAR-ship detection neural network via bidirectional coordinate attention and multi-resolution feature fusion".

[13] X. Ma, J. Cheng, A. Li, Y. Zhang, and Z. Lin, "AMANet: Advancing SAR ship detection with adaptive multi-hierarchical attention network".

[14] "SAR SHIP DETECTION DATASET SSDD" github.com/TianwenZhang0825/Official-SSDD.

[15] Tianwen Zhang, Xiaoling Zhang, Jianwei Li, Xiaowo Xu, Baoyou Wang, Xu Zhan, Yanqin Xu, Xiao Ke, Tianjiao Zeng, Hao Su, Israr Ahmad, Dece Pan, Chang Liu, Yue Zhou, Jun Shi, and Shunjun Wei, "SAR Ship Detection Dataset (SSDD): Official Release and Comprehensive Data Analysis," Remote Sens. 2021, 13(18), 3690.

[16] Yao Chen, Tao Duan, Changyuan Wang, Yuanyuan Zhang and Mo Huang, "End-to-End Ship Detection in SAR Images for Complex Scenes Based on Deep CNNs".

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun., "Deep Residual Learning for Image Recognition".

[18] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Ful, Alexander C. Berg, "SSD: Single Shot MultiBox Detector".

[19] "labelimg" github.com/qaprosoft/labelImg

[20] sgrvinod/a-PyTorch-Tutorial-to-Object-Detection

# A System to accurately detect and categorize Maritime Targets in Synthetic Aperture Radar Images.pdf

ORIGINALITY REPORT

| 9% | 6% | 6% | 2% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | academic-accelerator.com<br>Internet Source | 1% |
| 2 | www.mdpi.com<br>Internet Source | 1% |
| 3 | bibliostreet.ru<br>Internet Source | 1% |
| 4 | Chunjie Zhang, Peng Liu, Haipeng Wang, Yaqiu Jin. "A review of recent advance of ship detection in single-channel SAR images", Waves in Random and Complex Media, 2023<br>Publication | 1% |
| 5 | arxiv.org<br>Internet Source | 1% |
| 6 | Submitted to University of Hull<br>Student Paper | <1% |
| 7 | towardsdatascience.com<br>Internet Source | <1% |

| | | | |
|---|---|---|---|
| 8 | Submitted to American Public University System<br>Student Paper | <1 % |
| 9 | www.hindawi.com<br>Internet Source | <1 % |
| 10 | www.researchgate.net<br>Internet Source | <1 % |
| 11 | Jaewon Beom, Sukgyu Koh, Hyung Seok Nam, Wonshik Kim, Yoonjae Kim, Han Gil Seo, Byung-Mo Oh, Sun Gun Chung, Sungwan Kim. "Robotic Mirror Therapy System for Functional Recovery of Hemiplegic Arms", Journal of Visualized Experiments, 2016<br>Publication | <1 % |
| 12 | Submitted to Thadomal Shahani Engineering College<br>Student Paper | <1 % |
| 13 | Xiaozhen Ren, Yanwen Bai, Zihao Zhang, Wei Xu, Lulu Tan. "SEFRNet: A SAR Ship Target Detection Network with Effective Feature Representation", IEEE Sensors Journal, 2024<br>Publication | <1 % |
| 14 | "Image and Graphics Technologies and Applications", Springer Science and Business Media LLC, 2019<br>Publication | <1 % |

| | | |
|---|---|---|
| 15 | Modasshir, Md.. "Object Classification, Detection and Tracking in Challenging Underwater Environment", University of South Carolina, 2024<br>Publication | <1 % |
| 16 | Tao Xue, Jiayi Zhang, Wen Lv, Long Xi, Xiang Li. "HCA-RFLA: A SAR Remote Sensing Ship Detection Based on Hierarchical Collaborative Attention Method and Gaussian Receptive Field-Driven Label Assignment Strategy", Electronics, 2024<br>Publication | <1 % |
| 17 | Tianwen Zhang, Xiaoling Zhang, Jun Shi, Shunjun Wei. "HyperLi-Net: A hyper-light deep learning network for high-accurate and high-speed ship detection from synthetic aperture radar imagery", ISPRS Journal of Photogrammetry and Remote Sensing, 2020<br>Publication | <1 % |
| 18 | backend.orbit.dtu.dk<br>Internet Source | <1 % |
| 19 | dspace.iee.ihu.gr<br>Internet Source | <1 % |
| 20 | research.tue.nl<br>Internet Source | <1 % |
| 21 | "Artificial Intelligence Oceanography", Springer Science and Business Media LLC, | <1 % |

2023
Publication

| | | |
|---|---|---|
| 22 | Lentin Joseph, Amit Kumar Mondal. "Autonomous Driving and Advanced Driver-Assistance Systems (ADAS) - Applications, Development, Legal Issues, and Testing", CRC Press, 2021<br>Publication | <1 % |
| 23 | Pranjit Das, P. S. Ramapraba, K. Seethalakshmi, M. Anitha Mary, S. Karthick, Boopathi Sampath. "chapter 22 Sustainable Advanced Techniques for Enhancing the Image Process", IGI Global, 2024<br>Publication | <1 % |
| 24 | Prasad S. Thenkabail. "Land Resources Monitoring, Modeling, and Mapping with Remote Sensing", CRC Press, 2015<br>Publication | <1 % |
| 25 | Yao Chen, Tao Duan, Changyuan Wang, Yuanyuan Zhang, Mo Huang. "End-to-End Ship Detection in SAR Images for Complex Scenes Based on Deep CNNs", Journal of Sensors, 2021<br>Publication | <1 % |

| | | | |
|---|---|---|---|
| Exclude quotes | Off | Exclude matches | Off |
| Exclude bibliography | On | | |