SHIP DETECTION IN COSMO-SKYMED SAR IMAGERY USING A NOVEL CNN-BASED DETECTOR: A CASE STUDY FROM THE SUEZ CANAL

Tamer Saleh[1,2], Shimaa Holail[1], Xingxing Weng[3], Xiongwu Xiao[1], Gui-Song Xia[1,3,4,*]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University,
China (tamer.mohamed@feng.bu.edu.eg)
[2] Geomatics Engineering Department, Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt
[3] School of Computer Science, Wuhan University, China
[4] National Engineering Research Center for Multi-media Software, School of Computer Science and Institute of
Artificial Intelligence, Wuhan University, China (guisong.xia@whu.edu.cn)

KEY WORDS: Ship detection, COSMO-SkyMed SAR, Self-learning augmentation segmentation (SLAS), Deep learning, Attention-guided feature refinement (AFR), Suez Canal.

ABSTRACT:

The Suez Canal, strategically located as the shortest international sea route, plays a crucial role in facilitating the transportation of goods between Asia and Europe. However, the occurrence of traffic disruptions within the Canal poses a serious threat to global trade, as evidenced by the recent incident of the container ship Ever Given, which ran aground in the Suez Canal on March 23, 2021. This event led to a complete blockade of the Canal that lasted for six days, resulting in a fleet of ships waiting to pass through the Canal. This highlights the need to monitor the Canal to prevent similar disturbances in the future. In this paper, we propose a CNN-based attention-guided self-learning framework for ship detection from 3m high-resolution COSMO-SkyMed SAR imagery acquired in April 2021 via the Egyptian Suez Canal. We introduce a self-learning augmented segmentation (SLAS) technique to augment the dataset with new ship samples by pseudo-labeling an unlabeled dataset. We also present the Attention-guided Feature Refinement (AFR) module to extract more significant semantic features and contextual information, especially for ships of varying sizes in SAR images. Finally, the AFR module is fed into a Region Proposal Network (RPN) to generate a set of proposal anchors, which are later used in a Deep Detection Network (DDN) for ship classification and localization. Our experimental results demonstrate that the proposed method outperforms current state-of-the-art detection models in terms of detection accuracy, particularly in complex coastal scenes, with an overall accuracy of up to 87% mean average precision (mAP).

1. INTRODUCTION

The Suez Canal is a vital international sea route that has linked the Mediterranean and the Red Sea since 1869. Annually, nearly 20,000 vessels, including 2,500 tankers, cross the canal, representing 15% of the world's maritime freight traffic. However, this high volume of traffic also poses a risk of disrupting global trade in the marine environment (Lee and Wong, 2021). The recent grounding of the Ever Given, a large container ship passing through the southern part of the canal on its way to Rotterdam on March 23, 2021, exemplified this risk (Forti et al., 2021), as shown in Figure 1. As one of the busiest trade routes globally, this blockage had a considerable negative impact on global trade, highlighting the need for effective monitoring and management of the channel to prevent similar disturbances in the future or, at least, reduce marine anomalies. Therefore, it is crucial to spot ships in the channel to ensure the safe and smooth operation of the waterway and manage fisheries, wrecks, and other potential risks. In recent years, synthetic aperture radar (SAR) images have become a popular tool for ship detection, as they offer imaging capabilities that are available 24/7, regardless of weather or lighting conditions (Maître, 2013); (Yoshida et al., 2021). The SAR beam interacts with the physical surface of the water, allowing it to detect physical objects such as waves and ships that differ from the surface of the water. Of the various search and rescue systems available, the Cosmo-SkyMed SAR (CSK-SAR) constellation stands out because of its high resolution and high repetition time (few hours). This

makes it particularly suitable for monitoring high traffic areas and detecting ships within channels. However, identifying vessels in SAR images is a challenging task due to the complex marine background. Figure 2 clearly illustrates typical challenges in detecting ships using high-resolution SAR images, including lighthouses (A), buildings (B), gantry cranes (C), side lobes (D), ghosts caused by azimuth ambiguity (E), sea clutter (F), and beach clutter (G). These appearances can be so similar to that of the vessel in Figure 2(h) that distinguishing them can be a counter intuitive challenge even for experts. Thus, they have to be mitigated before the data can be applied to machine learning algorithms so as not to lead to false alarms.

Recently, researchers have increasingly employed machine learning techniques to develop methods for detecting ships in SAR images. (Wu et al., 2011) presented a multi-scale processing algorithm that optimizes target pixels and suppresses speckle and background clutter. The algorithm's efficacy was tested on a Cosmo-Skymed image with a 3-meter pixel spacing, demonstrating its ability to detect ships effectively. (Liu et al., 2017) introduced a Sea-Land Segmentation based CNN (SLC-CNN) that combines corner features and saliency. The model underwent testing on high-resolution images from the TerraSar-X and ALOS PALSAR datasets. (Huang et al., 2022) introduced WA-CNN, a novel CNN-based method for ship detection in SAR images, which employs wavelets and an attention mechanism. The U-Net architecture was utilized to develop the network, reducing depth and enhancing complexity. The approach was evaluated on two public SAR image datasets and

---

* Corresponding author

Figure 1. A striking view captured from a Sentinel-1 satellite on March 25, 2021, displays a fleet of ships at a standstill, patiently waiting to pass through the southern entrance of the Suez Canal. Upon magnifying the image, both COSMO-SkyMed SAR and optical high-resolution imageries reveal the colossal Ever Given container ship, still stranded in the canal's narrow passage. The credits for this awe-inspiring composition go to Copernicus Sentinel 1 data©ESA, CC BY-SA 3.0 IGO; COSMO-SkyMed image©ASI, processed and distributed by e-GEOS; and Satellite image©2021 Maxar Technologies.
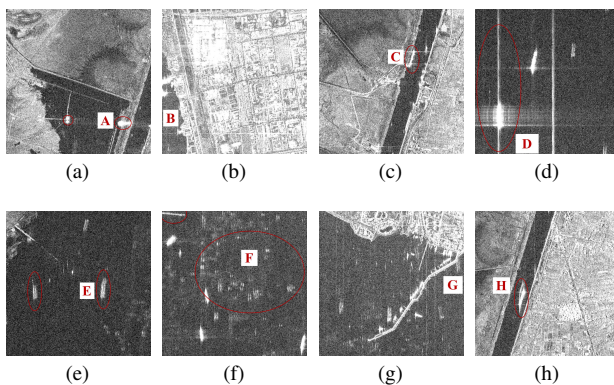

Figure 2. Samples typical challenges in ship detection from high-resolution COSMO-SkyMed SAR (CSK-SAR) images containing land, ships, and marine clutter. COSMO-SkyMed®Products©ASI-Italian Space Agency-2021. All rights reserved.

yielded promising results. Despite these promising outcomes, detecting ships in SAR images remains a challenging task, particularly in complex coastal scenes. These approaches also encounter other challenges, such as the need for large annotated datasets and the selection of appropriate hyper-parameters. Recent advancements in deep learning methods have significantly improved ship detection in SAR images. (Wang et al., 2018) introduced an enhanced Single-Shot Detector (SSD) model that can simultaneously detect ships and estimate their orientation angles. Attentional modules were incorporated to enhance the model's performance. Similarly, (Miao et al., 2022) enhanced

the RetinaNet model for vessel detection by introducing attention modules and adjusting aspect ratios using the k-means clustering algorithm. These modifications improved the accuracy and effectiveness of the model. (Li et al., 2017) proposed an improved Faster R-CNN-based technique for ship detection, utilizing transfer learning and feature fusion. Their approach demonstrated robustness and effectiveness when evaluated on high-resolution images from various datasets. (Li et al., 2022) further improved the Cascade-R-CNN method by incorporating the Swin transformer, leading to enhanced performance and accuracy. Despite these advancements, challenges persist in detecting ships in complex coastal scenes, and a significant resolution gap between beach and marine scenes presents an additional obstacle in ship detection.

This paper proposes a novel framework for ship detection from high-resolution COSMO-Skymed SAR (CSK-SAR) datasets acquired in April 2021 via the Egyptian Suez Canal. Our approach employs the open-source Faster R-CNN model as the primary attention-guided self-learning design framework. Specifically, we propose a Self-Learning Augmented Segmentation (SLAS) technique to augment the dataset with new ship samples by pseudo-labeling an unlabeled dataset. Furthermore, we propose an Attention-guided Feature Refinement (AFR) module to enable the convolutional layers to extract more meaningful semantic features, especially about coastal ships of different sizes in SAR images, by leveraging global contextual information. The output of the AFR module is then fed into a Region Proposal Network (RPN) to generate a set of proposal anchors that are later utilized in a Deeply Detection Network (DDN) for classification and localization. Our experimental results demonstrate that the proposed method outperforms the latest ship detection models in terms of detection accuracy and efficiency, particularly in complex coastal scenes. This indicates the potential of our approach for real-world applications. The remaining parts of this paper are organized as follows. In Section 2, the proposed method is listed. The CSK-SAR dataset, training details, evaluation criteria, and results obtained are described in Section 3. Finally, some conclusions are drawn in Section 4.

2. PROPOSED METHODOLOGY

2.1 Overall Scheme of Network Structure

The current state-of-the-art object detection models can be classified into two categories: single-stage models and two-stage models. Single-stage detectors are known for their fast inference speeds, which are particularly important in applications that require low latency, such as video detection. On the other hand, two-stage detectors are characterized by their high accuracy in locating and recognizing objects within an image. Considering the challenges involved in detecting ships in SAR images and the significant number of false alarms generated, we opted for the Faster R-CNN two-stage model as the basic design framework for accurate in-image ship recognition (Ren et al., 2015). The Faster R-CNN model's ability to learn features at different scales is particularly important in detecting ships of varying sizes and orientations. For comparison purposes, we chose three other representative methods - Cascade R-CNN (Cai and Vasconcelos, 2018), RetinaNet (Lin et al., 2017b), and SSD (Liu et al., 2016) in our work. The proposed method's network structure is detailed in Figure 3 and can be divided into five parts: the self-learning augmentation segmentation (SLAS) part, the feature extraction network (FEN) part,
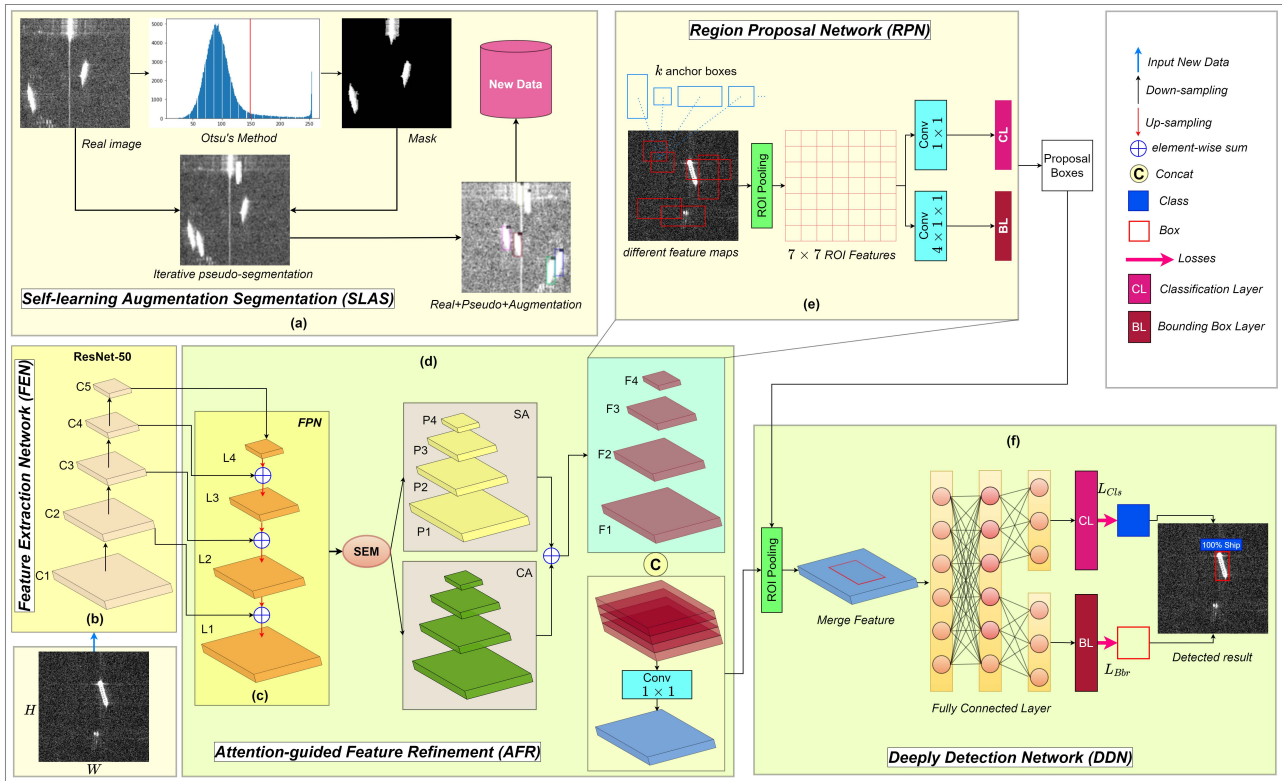
Figure 3. Overview of our proposed network for ship detection from SAR images.

the attention-guided feature refinement (AFR) part, the region proposal network (RPN) part, and the deeply detection network (DDN) part. We present a detailed explanation of each of these five parts in the following sections.

2.2 Self-Learning Augmentation Segmentation

In this section, we present self-learning augmentation segmentation (SLAS), which is inspired by (Ghiasi et al., 2021) and involves augmenting the dataset with new ship samples through self-learning and iterative pseudo-segmentation. We generated masks for each ship in the image using the Otsu method (Yu et al., 2010) on the COSMO-SkyMed SAR (CSK-SAR) dataset. This enabled us to identify all the pixels that belonged to each ship in the image. From our dataset, we selected 1,512 annotated images, some of which contained only background and others contained ships. We identified a total of 136 segmentation instances of ships in these images. To augment the data, we applied random rotations, brightness adjustments, contrast enhancements, lighting modifications, and saturation changes to the 1,512 image dataset. Furthermore, the iterative pseudo-segmentation process resulted in two or three additional copies of each ship instance segmentation. We added these additional copies to a new dataset containing labeled ships and not only pure background. After this initial processing, we used the new augmented dataset during iterative pseudo-segmentation to train our proposed network.

2.3 Feature Extraction Network

During the feature extraction network (FEN) phase, we slice the new augmented dataset into smaller patches of size 500 × 500 pixels, filtering out patches that do not contain ships, randomly splitting the remaining patches into training, validation, and testing sets. These patches are then fed into a deep convolutional neural network (CNN) to obtain feature maps at various

scales. To construct the FEN, we utilized the ResNet-50 (He et al., 2016) architecture with Faster R-CNN as the backbone network. This design plays a crucial role in reducing the size of the network while increasing its depth. Instead of stacking convolutional layers directly, ResNet-50 connects these layers to fit the residual mapping. Specifically, the input SAR image x is processed with two weight convolutional layers and an activation function to yield $F(x)$. This is then added to x to obtain $H(x)$, which is further processed to yield the final output z, as depicted in Figure 4. As shown in Figure 3(b), the FEN consists of five hierarchical convolutional layers $C1 - C5$, which generate deep feature maps from the input SAR images. The lower layers ($C1$ and $C2$) have higher spatial resolution but contain less semantic information, while higher layers ($C4$ and $C5$) have more abstract, semantic information but lower spatial resolution. As a result, the location of the ships in the upper layers is coarser. To address this, we adopt the idea of fine-tuning attention-guided feature refinement (AFR), which integrates the feature information of all convolutional layers to make full use of semantic and spatial information.
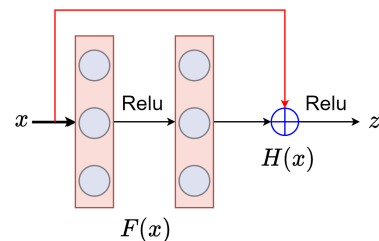

Figure 4. The shortcut connection of ResNet.

2.4 Attention-guided Feature Refinement

In our proposed attention-guided feature refinement (AFR) fusion network, we first adopt a feature pyramid network (FPN)

(Lin et al., 2017a) to generate multi-scale feature maps that capture rich semantic information at different levels. The FPN architecture involves both bottom-up and top-down processes, as depicted in Figure 3(b-c). The resulting feature maps are then used by both ResNet-50 and FPN to make detectors suitable for detecting shore ships of different sizes in SAR images. Next, we introduce the AFR module, which consists of four feature fusions ($F4$, $F3$, $F2$, and $F1$), as shown in Figure 3(d). First, we down-sample the last convolutional layer $C5$ and combine it with the $C4$ layer using an element-wise sum operation, to obtain the feature map $L4$. This process is repeated until we obtain the most accurate feature map. However, since SAR images often contain complex background environment information, it is essential to direct the network's attention to the features that are most distinguishable for the current detection task, namely ships. To this end, we propose to use the Squeeze-Excitation module (SEM) (Hu et al., 2018) to encode the feature maps and provide a weight for each channel of the feature maps.

2.5 Region Proposal Network

In this section, we present the use of a Region Proposal Network (RPN) in combination with each feature map fusion Fi to achieve high performance in the detection of ships in SAR images, as depicted in Figure 3(e). The RPN generates a set of reference boxes, or anchors, called region proposals. To cover ship targets with different sizes for each Fi layer, we used four different scales of anchors, denoted as $\text{Scale}_i = \{64 \times 64, 128 \times 128, 256 \times 256, 512 \times 512\}$ where $i \in \{1, 2, 3, 4\}$. In addition, we used an aspect ratio of $\{1 : 1, 1 : 2, 2 : 1\}$, resulting in a total of 12 (4 scales in 3 aspect ratios) anchors for each i anchor. To generate fixed-dimension 7×7 Region of Interest (RoI) features from the anchor proposals, we adopted a RoI pooling layer. Subsequently, the anchors were sent to both the ship classification layer (CL) and the box regression layer (BL) through two convolutional layers. The CL has a $2K$ output of the object probability estimation for each proposal, while the BL has a $4K$ output of the bin coding coordinates (where $K = 12$).

2.6 Deeply Detection Network

In Figure 3(f), the deeply detection network (DDN) represents the final stage of our proposed framework. This stage leverages the fusion of enhanced features extracted from the AFR and proposal anchors generated by RPN as inputs. To enhance the semantic information pertaining to small-sized ships, we combine the features generated by the RoI pooling layer with the AFR-enhanced features. The output is then fed back to the fully connected layers with the sigmoidal activation function to obtain the final detection result. From the fully connected layers, the detection result branches into two simultaneous branches: the classification layer (CL) branch and the bounding box regression branch (BL). The CL confidently predicts the corresponding prior anchors for each pixel on feature maps at different scales. This confidence determines the probability that the anchor belongs to the class of ships. Meanwhile, the BL outputs the offsets of the coordinates between the anchor belonging to a ship and the ground-truth bounding box of that ship. In our work, we adopt a double loss approach from CL and BL to detect each layer, as described in Equation 1. Specifically, the CL introduces the sigmoid activation function to obtain the detected probability P and ground truth Y. We then calculate the classification loss indicated by $\mathcal{L}_{Cls}$. We use λ as the balancing parameter, where $Y \geq 1$ indicates that the background is meaningless for training the BL, and $\mathcal{L}_{Bbr}$ indicates the loss of

the bounding box regression. In $\mathcal{L}_{Bbr}$, we predict four offsets for anchor box i, which are $A_i = (a_i^x, a_i^y, a_i^w, a_i^h)$, where a_i^x and a_i^y are the top-left coordinates of the predicted area, and a_i^w and a_i^h refer to the width and height of the projected area. If the predicted area of anchor box $\widehat{A}$ has the highest IoU with the ground truth box A^*, we assign it a positive sign $Y_i \geq 1$. Conversely, if the IoU ratio of the predicted area's box is less than 0.3 for all ground truth boxes, we assign a negative label $Y_i = 0$ to it, and then ignore the remaining regions. The IoU ratio is defined as in Equation 2.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Cls}}(P, Y) + \lambda[Y \geq 1]\mathcal{L}_{\text{Bbr}}\left(A^*, \widehat{A}\right)$$
$$\mathcal{L}_{\text{Cls}}(P, Y) = -\log\left(P_Y\right)$$
$$\mathcal{L}_{\text{Bbr}}\left(A^*, \widehat{A}\right) = \text{Smooth}_{L1}\left(A^* - \widehat{A}\right) \quad (1)$$
$$\text{Smooth}_{L1}(z) = \begin{cases} 0.5\left(z^2\right), & \text{if } |z| < 1 \\ |z| - 0.5, & \text{Otherwise} \end{cases}$$

$$\text{IoU}_{\text{bbox}} = \frac{\text{area}\left(\widehat{A} \cap A^*\right)}{\text{area}\left(\widehat{A} \cup A^*\right)} \quad (2)$$

Where $\text{area}\left(\widehat{A} \cup A^*\right)$ is the union of the predicted area's box and the ground truth's box, and $\text{area}\left(\widehat{A} \cap A^*\right)$ is their intersection.

3. EXPERIMENTAL RESULTS

3.1 Datasets

In this study, X-band SAR images from the COSMO-SkyMed system were utilized. COSMO-SkyMed is a constellation of four SAR satellites developed by Agenzia Spaziale Italiana (ASI), which enables targeting of the same location on Earth within a single day. Two scenes of COSMO-SkyMed StripMap SAR (CSK-SAR) data with horizontal-horizontal (HH) and vertical-vertical (VV) polarizations were used in this research. The datasets used were level 1A (L1A) HIMAGE Single-look Complex Slant (SCS) products, provided by ASI as part of the COSMO-SkyMed project (Open Call Id 797). Table 1 presents detailed parameters of the test image datasets, while Figure 5(e) shows the test images acquired from the COSMO-SkyMed satellite. The original CSK-SAR datasets were calibrated, converted to floating-point numbers in dB, and exported in a tagged image file format TIFF, with each image containing 18000×15000 pixels. The entire image was divided into 2160 sub-images of size 500×500 pixels. Of these sub-images, 1512 were used as the training set, and 648 were used as the testing set. The test images were further divided into 486 offshore sub-images (test offshore) and 162 inshore sub-images (test inshore). The annotated dataset was created through a manual annotation process using the LabelMe Toolbox.

3.2 Training Details

To implement our proposed method, we utilized Detectron2, a flexible and powerful toolkit for re-implementing existing object detection models (Wu et al., 2019). We used four open-

Type	SCS−B(L1A)	SCS−B(L1A)
Satellite Mission	CSKS2	CSKS4
Acquisition Mode	StripMap	StripMap
Polarization	VV	HH
Resolution	3m	3m
Product Format	HDF5	HDF5
Acquisition Date	09 April, 2021	08 April, 2021
Time	15:22:59.71	15:23:05.27
Look Side	Right	Right
# pixels in X	15568	15568
# pixels in Y	17893	17893
Orbit	72145	56381
Swath Extent	50 x 40 km	50 x 40 km

Table 1. Detailed image parameters of the X-band SAR images from COSMO-SkyMed SAR Datasets.
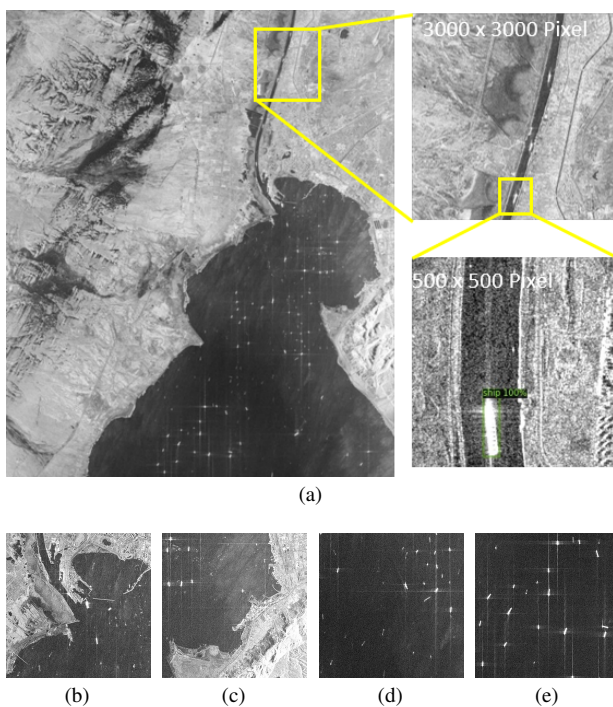

(a)

(b)　　(c)　　(d)　　(e)

Figure 5. An example of large-scale image and sub-images from COSMO-SkyMed SAR imagery with 3000×3000 pixels, sub-figure (a) shows the SAR imagery acquisition area of our dataset and distributed ships in the Suez Canal. The sub-figures (b) and (c) represents the inshore scenes. Offshore scenes with multiple ships, as displayed in (d) and (e).
COSMO-SkyMed®Products©ASI-Italian Space Agency-2021.
All rights reserved.

source deep learning detectors, including Faster R-CNN, Cascade R-CNN, SSD, and RetinaNet, and employed ResNet-50+FPN pre-trained on the COCO dataset as the backbone for these models. To improve the accuracy of location and segmentation, we resized all samples of CSK-SAR imagery to 512×512 pixels for both network training and testing. We trained all detectors using GPUs and completed the training in 52 epochs via stochastic gradient descent (SGD) as the optimizer. The momentum, weight decay, learning rate, and batch size were set to 0.9, 0.0001, 0.02, and 4, respectively. We performed all experiments on our dataset using a virtual machine desktop with a 64-bit Windows 10 Pro operating system. The software configuration included the Python programming language, PyTorch 1.6.0, CUDA 10.1, and cuDNN 7.6.1. Addi-

tionally, the hardware capabilities comprised an NVIDIA GRID RTX8000-8Q with 8GB memory, an Intel(R) Xeon(R) CPU E5-2687W v4 @ 3.00GHz, and 32.0 GB RAM.

3.3 Evaluation Indices

To evaluate the performance of the proposed model against the state-of-the-art models, the following evaluation metrics were considered for the testing set: Detection Probability (Pd), False Alarm (Pf), Missed Detection (Pm), Recall, Precision, Mean Average Precision (mAP), F1-score, True Positive (TP), Ground Truth (GT), False Positive (FP), and False Negative (FN).

$$P_d = \frac{TP}{GT}; P_f = \frac{FP}{TP+FP}; P_m = \frac{FN}{GT}$$

$$R = \frac{TP}{TP+FN}$$

$$P = \int_0^1 P(R)\, dR \qquad (3)$$

$$F1 = 2\frac{P*R}{P+R}$$

where
P = precision
R = recall
$P(R)$ = precision-recall curve

3.4 Results and Analysis

In Table 2, we have shown the ship detection statistics in inshore and offshore scenes resulting from the bounding box AP on the testing set of CSK-SAR datasets. One can observe that AP_{50} of state-of-the-art detectors was above 82% for detecting the offshore scenes. While detecting the inshore scenes, AP_{50} of state-of-the-art detectors was above 75%. Hence, the detection accuracy has dropped significantly for inshore scenes by 7%. Moreover, Faster R-CNN with SLAS adoption leads better in bounding box AP than the state-of-the-art detectors. From Table 3, one can observe that on the offshore+inshore scenes, SLAS enhanced SSD, RetinaNet, Cascade R-CNN, and Faster R-CNN by 5%, 3%, 4%, and 3% mAP, respectively. As for validation and testing inference, the precision-recall curve (PR curve) and IoU between predictions and ground truth obtained from inference on inshore and offshore scenes using Faster-RCNN algorithm with SLAS adoption are showing in Figure 7(a-b). Learning curves for training, validation loss, and validation mAP appears in Figure 7(c-d). Figure 8 displays the outcomes of a small subset of the CSK-SAR dataset utilized to evaluate the detection capability of our model. Incorporating the proposed SLAS with ResNet50+FPN and Faster R-CNN, we observe that the detection results exhibit commendable ship detection performance. To study the ship detection ability of detectors to complex CSK-SAR scenes intuitively, we have selected 4 representative scenes as an example in the test set and use the Faster-RCNN algorithm with SLAS adoption to detect ships. At the same time, we took some auxiliary means such as optical remote sensing images that have the same imaging date for CSK-SAR images from Google Earth to determine the potential interference between similar features of targets in the Canal. The visual results between predictions and ground truth on test set without/with the back-scattering appears in Figure 6. The number in the predicted bounding box represents the confidence of the detection box and was filtered under the 0.7 confidence coefficient. As a result, most of small and large ships were detected with correct rectangles on the offshore scenes.

Detector's type	Method	Backbone	Scenes	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
One-stage	SSD	ResNet-50+FPN	Inshore	0.53	0.78	0.66	0	0.61	0.47
			Offshore	0.54	0.79	0.59	0.17	0.60	0.49
		ResNet-50+FPN+SLAS	Inshore	0.53	0.79	0.65	0	0.60	0.45
			Offshore	0.55	0.81	0.59	0.22	0.61	0.51
	RetinaNet	ResNet-50+FPN	Inshore	0.54	0.80	0.68	0	0.66	0.50
			Offshore	0.53	0.87	0.60	0.19	0.62	0.56
		ResNet-50+FPN+SLAS	Inshore	0.54	0.82	0.56	0	0.63	0.40
			Offshore	0.54	0.88	0.60	0.23	0.60	0.58
Two-stage	Cascade R-CNN	ResNet-50+FPN	Inshore	0.54	0.75	0.75	0	0.62	0.45
			Offshore	0.56	0.82	0.64	0.18	0.62	0.54
		ResNet-50+FPN+SLAS	Inshore	0.53	0.77	0.77	0	0.61	0.45
			Offshore	0.56	0.85	0.68	0.24	0.62	0.54
	Faster R-CNN	ResNet-50+FPN	Inshore	0.57	0.76	0.67	0	0.67	0.40
			Offshore	0.54	0.88	0.64	0.15	0.62	0.56
		ResNet-50+FPN+SLAS	Inshore	0.58	**0.83**	0.70	0	0.69	0.45
			Offshore	0.58	**0.90**	0.63	0.26	0.63	0.64

Table 2. Ship detection statistics generated by bounding box AP on test set of CSK-SAR dataset.

Inshore + Offshore

Model	SLAS	Pf	Pm	Recall	Precision	mAP	F1-score
SSD	No	0.49	0.09	0.90	0.51	0.82	0.64
	Yes	0.46	0.07	0.94	0.57	0.87	0.71
RetinaNet	No	0.47	0.05	0.94	0.52	0.85	0.67
	Yes	0.45	0.05	0.95	0.54	0.88	0.69
Cascade R-CNN	No	0.15	0.17	0.82	0.84	0.79	0.83
	Yes	0.21	0.13	0.87	0.78	0.83	0.82
Faster R-CNN	No	0.23	0.09	0.90	0.76	0.84	0.82
	Yes	0.22	0.08	0.92	0.77	0.87	0.84

Inshore

Model	SLAS	Pf	Pm	Recall	Precision	mAP	F1-score
SSD	No	0.60	0.20	0.88	0.46	0.71	0.55
	Yes	0.52	0.22	0.89	0.54	0.79	0.68
RetinaNet	No	0.55	0.14	0.85	0.44	0.74	0.58
	Yes	0.48	0.14	0.86	0.52	0.77	0.65
Cascade R-CNN	No	0.21	0.21	0.78	0.78	0.69	0.78
	Yes	0.31	0.21	0.79	0.69	0.70	0.73
Faster R-CNN	No	0.26	0.21	0.78	0.73	0.69	0.75
	Yes	0.25	0.14	0.86	0.75	0.76	0.80

Offshore

Model	SLAS	Pf	Pm	Recall	Precision	mAP	F1-score
SSD	No	0.47	0.07	0.94	0.52	0.84	0.70
	Yes	0.48	0.08	0.96	0.56	0.87	0.73
RetinaNet	No	0.45	0.03	0.97	0.54	0.86	0.69
	Yes	0.45	0.03	0.97	0.55	0.89	0.70
Cascade R-CNN	No	0.13	0.16	0.83	0.86	0.80	0.85
	Yes	0.19	0.11	0.89	0.80	0.84	0.84
Faster R-CNN	No	0.23	0.07	0.93	0.77	0.86	0.84
	Yes	0.21	0.06	0.93	0.78	0.87	0.85

Table 3. Comparison between different baselines with or/without SLAS results on test inference.
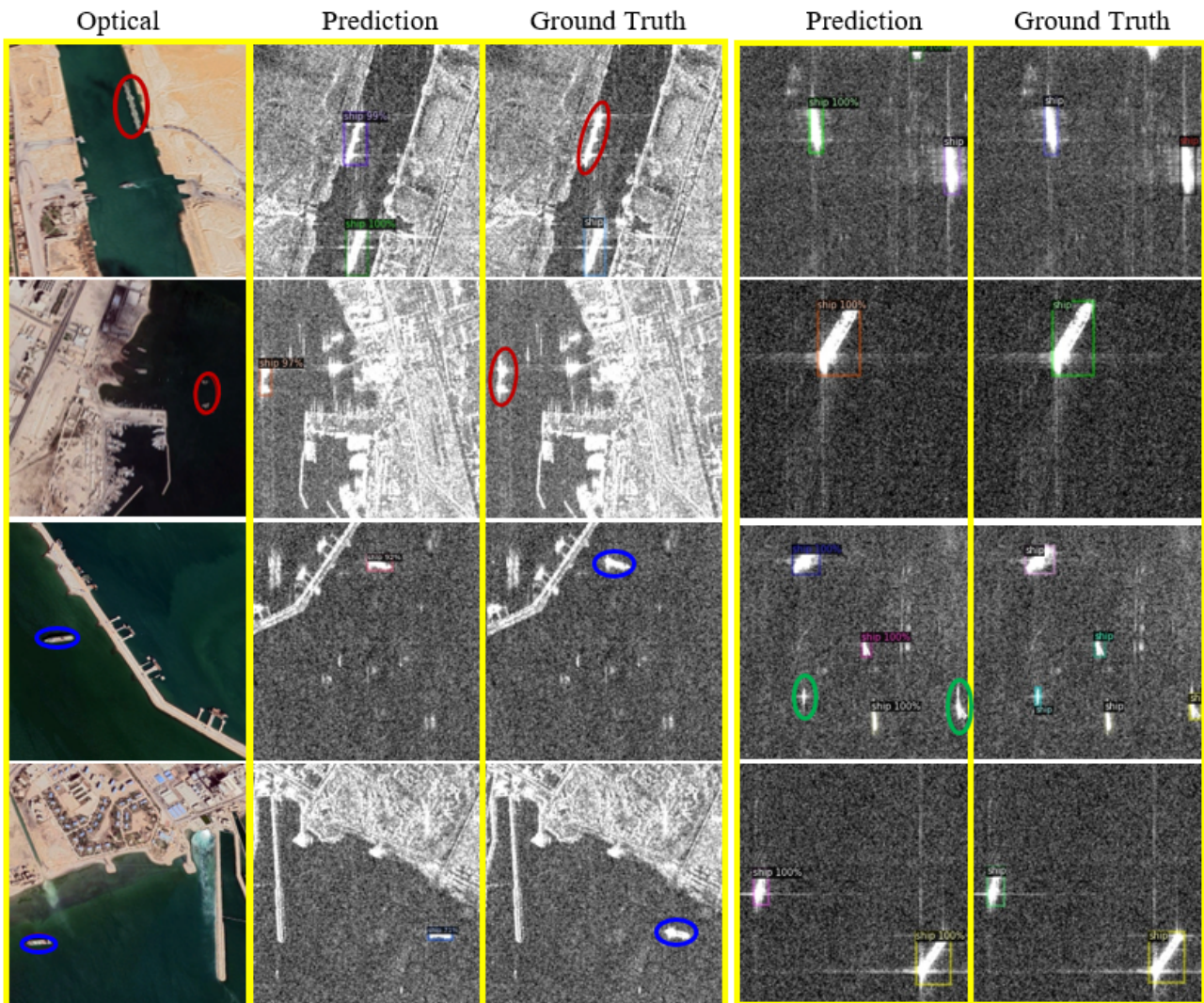
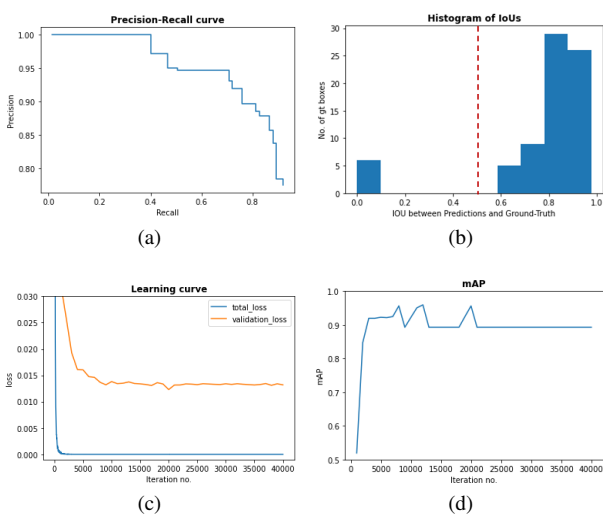Figure 6. Comparison between predictions and ground truth on test set without/with back-scattering.

Figure 7. Learning curves for training and inference for testing, the detection PR curve displayed in sub-figure (a). The IOU between predictions and ground truth, as showed in (b). The Validation loss curve, as displayed in (c). The Validation mAP, as showed in (d).
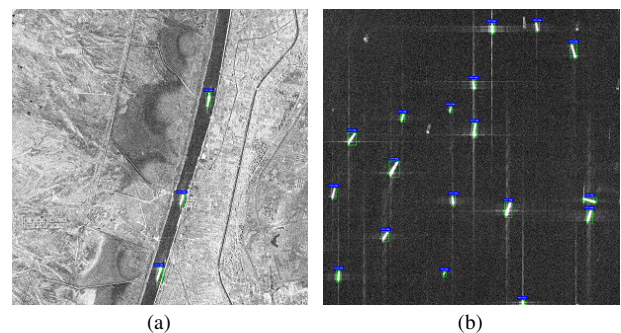

Figure 8. The outcomes of a small subset of the CSK-SAR dataset by Faster R-CNN+ResNet50+FPN+SLAS.

4. CONCLUSIONS

In this paper, we propose an improved attention-guided self-learning Faster R-CNN-based framework for detecting ships in-shore and offshore from COSMO-SkyMed SAR datasets across the Egyptian Suez Canal. Our experiments demonstrate that the proposed Self-Learning Augmented Segmentation (SLAS) technique effectively augments data by placing pseudo-labeled ships on an unlabeled dataset, thereby significantly improv-

ing model performance. Additionally, the proposed Attention-guided Feature Refinement (AFR) module enables convolutional layers to extract more meaningful semantic features, particularly around coastal ships of varying sizes in SAR images, by leveraging global contextual information. This feature refinement enhances the model's detection ability. Our model outperforms other competing detectors in AP bounding box and demonstrates superior ship detection ability from SAR images, as evidenced by our experiments. However, the current model has limitations in mistakenly identifying non-ship targets at sea and using a horizontal rectangular bounding box to mark ship targets, resulting in poor performance when ship targets are in close proximity inshore. To improve the accuracy of ship detection and reduce the performance gap between inland and offshore scenes, future research should consider land and sea segmentation before detection, along with the use of more diverse SAR datasets.

ACKNOWLEDGMENT

REFERENCES

Cai, Z., Vasconcelos, N., 2018. Cascade r-cnn: Delving into high quality object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6154–6162.

Forti, N., d'Afflisio, E., Braca, P., Millefiori, L. M., Willett, P., Carniel, S., 2021. Maritime anomaly detection in a real-world scenario: Ever Given grounding in the Suez Canal. *IEEE Transactions on Intelligent Transportation Systems*, 23(8), 13904–13910.

Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E. D., Le, Q. V., Zoph, B., 2021. Simple copy-paste is a strong data augmentation method for instance segmentation. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2918–2928.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141.

Huang, S., Pu, X., Zhan, X., Zhang, Y., Dong, Z., Huang, J., 2022. SAR ship target detection method based on CNN structure with wavelet and attention mechanism. *Plos one*, 17(6), e0265599.

Lee, J. M.-y., Wong, E. Y.-c., 2021. Suez canal blockage: an analysis of legal impact, risks and liabilities to the global supply chain. *MATEC web of conferences*, 339, EDP Sciences, 01019.

Li, J., Qu, C., Shao, J., 2017. Ship detection in sar images based on an improved faster r-cnn. *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)*, IEEE, 1–6.

Li, K., Zhang, M., Xu, M., Tang, R., Wang, L., Wang, H., 2022. Ship detection in SAR images based on feature enhancement Swin transformer and adjacent feature fusion. *Remote Sensing*, 14(13), 3186.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017a. Feature pyramid networks for object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017b. Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A. C., 2016. Ssd: Single shot multibox detector. *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, Springer, 21–37.

Liu, Y., Zhang, M.-h., Xu, P., Guo, Z.-w., 2017. Sar ship detection using sea-land segmentation-based convolutional neural network. *2017 international workshop on remote sensing with intelligent processing (RSIP)*, IEEE, 1–4.

Maître, H., 2013. *Processing of synthetic aperture radar (SAR) images*. John Wiley & Sons.

Miao, T., Zeng, H., Yang, W., Chu, B., Zou, F., Ren, W., Chen, J., 2022. An improved lightweight retinaNet for ship detection in SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 4667–4679.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

Wang, J., Lu, C., Jiang, W., 2018. Simultaneous ship detection and orientation estimation in SAR images based on attention module and angle regression. *Sensors*, 18(9), 2851.

Wu, F., Wang, C., Zhang, H., Zhang, B., 2011. Vessel detection method for high resolution cosmo-skymed sar imagery. *2011 IEEE International Geoscience and Remote Sensing Symposium*, Ieee, 1798–1801.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., Girshick, R., 2019. Detectron2. `https://github.com/facebookresearch/detectron2`.

Yoshida, T., Ouchi, K., Yang, C.-S., 2021. Application of MA-ATI SAR for estimating the direction of moving water surface currents in Pi-SAR2 images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 2724–2730.

Yu, C., Dian-Ren, C., Yang, L., Lei, C., 2010. Otsu's thresholding method based on gray level-gradient two-dimensional histogram. *2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010)*, 3, IEEE, 282–285.