Automated Ship Detection and Characterization in Sentinel-2 Images: A Comprehensive Approach

Bou-Laouz Moujahid, Vadaine Rodolphe, Hajduch Guillaume, Ronan Fablet

Automated Ship Detection and Characterization in Sentinel-2 Images: A Comprehensive Approach

Bou-Laouz Moujahid[1], Vadaine Rodolphe[2], Hajduch Guillaume[2], and Fablet Ronan[1]

[1]IMT Atlantique, Plouzané France
[2]CLS, Collecte Localisation Satellites, Plouzané France

***Abstract*—The automatic detection and characterization of ships in optical remote sensing images is a key challenge for maritime surveillance applications. This paper presents an automated system specifically designed for ship detection in medium-resolution Sentinel-2 images. The proposed approach relies on a deep learning model trained on a dataset comprising over 6000 annotated Sentinel-2 images. It achieves a detection rate of 93%, with an average of 2.1 to 3.9 false alarms per Sentinel-2 image. Besides the detection task, it also addresses the estimation of ship lengths as well as ship headings. It yields a mean error of 15.36m $\pm$ 19.57m for ship lengths, and estimates ship headings with an accuracy of 93%. This contribution significantly enhances the performance of ship detection and characterization systems in optical remote sensing imagery.**

Index Terms: Deep Neural Network, Sentinel-2 Images, Ship Detection, Ship Characterization

I. Introduction

Spaceborne remote sensing imagery conveys invaluable information for the monitoring of maritime activities, especially the maritime traffic. This is of critical importance for both surveillance and defense issues [1]. We may cite among others the monitoring of maritime borders, the identification of illegal maritime behaviours, the fight against illegal fishing and smuggling activities, search and rescue operations...

We can distinguish two main categories of satellite imagery for maritime surveillance topics: Synthetic Aperture Radar (SAR) imagery [2] and optical imagery [3]. Currently, SAR imagery is more widespread due to its applicability both day and night and in all weather conditions (e.g., cloudy conditions). Ship detection in SAR images [4] relies on relatively simple low-level image processing schemes for ships which possess a metallic structure highly responsive to radar signals. This may lead to detection ambiguities along the seashore, where other metallic structures (buoys, pontoons, etc.) cause high-amplitude patterns in SAR images [5], as well as in the case of sea clutter [6]. Additionally, inflatable and wooden boats like Zodiacs and plastic boats can hardly be identified in SAR images. By contrast, optical imagery offers a practical alternative for ship detection, increasing not only detection capabilities in complex environments but also providing the ability to detect all types of ships.

In recent years, ship detection from optical imagery has seen increased research effort, with a focus on deep learning methods [7]. Most studies focus on ship detection from very-high-resolution (VHR) (0.3 m resolution) [8] and high-resolution (HR) ($0.4 - 2$ m resolution) [9] imagery. However, these satellites images are typically accessible exclusively through tasking modes, resulting in high acquisition costs and a limited interest for continuous monitoring and surveillance tasks. By contrast, medium-resolution (MR) optical imagery (typically, a 10 m resolution), as deployed on Sentinel-2, delivers freely available remote sensing images for numerous locations on Earth with a revisit time ranging between 5 and 10 days. This seems particularly adapted to maritime surveillance tasks. Yet, only few studies [7] have addressed the automated detection and characterization of ships in MR optical satellite imagery.

Here, we address these challenging issues and present a deep learning approach. We first collect a representative groundtruthed dataset comprising more than 12000 ship examplars from 6000 Sentinel-2 images. Our multi-task deep learning scheme relies on a Faster R-CNN. Our numerical experiments explore data splitting strategies during the training phase to account for class imbalance. We also asses the impact of the neural backbone of the Faster R-CNN architecture. Overall, we report state-of-the-art performance with a detection rate above 93% including for small ships with a ship length below 20 m. These results support the relevance MR optical imagery for maritime monitoring besides HR optical and SAR imagery.

This paper is organized as follows. In Section II, we provide an overview of related work and analyze their drawbacks. The details of the proposed approach are presented in Section III. Section IV substantiates the relevance of our proposed methods through experiments. Discussion is covered in Section V, and the conclusion is presented in Section VI.

II. Related work and Motivation

Recent review papers [7], [10] provide surveys on ship detection and characterization in optical satellite images. They compile over a hundred research articles dating back to 1978, providing a comprehensive overview of the subject. The majority of these studies have investigated ship detection using High-Resolution (HR) and Very High-Resolution (VHR) images, employing deep learning approaches such as Faster R-CNN [11], YOLO [12] and U-net [13]. However, our specific challenge is different from working with high or very high-resolution (<5 m) images. In these cases, identifying ships among other objects is relatively straightforward due to

their visually distinct features. As a result, the challenge of ship detection in high-resolution images primarily revolves around maximizing the detection rate and closely aligns with the established problem of object detection in computer vision, involving issues such as adjacent and small object detection [11].

In medium-resolution imagery ($10\,\mathrm{m}$ in our case), as illustrated in Figure 1, ships, especially smaller ones of 50-meter-long or below, the visual detection and characterization may be complex: for instance, a large ship can resemble parts of a platform; a small island may have a topology similar to a ship; a very small ship might appear almost identical to a bright spot on rough waters, or even be mistaken for a small cloud. As a result, a key challenge is to maximize the detection rate while minimizing false alarms. This may question how above cited studies for HR images apply to MR optical imagery.

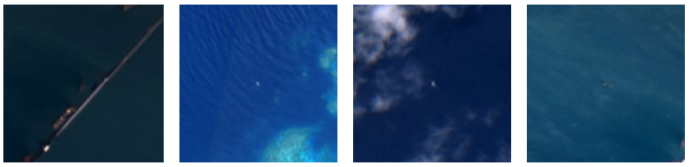


Fig. 1: Illustration of complex ship images, presenting challenges in visual detection and characterization.

Only a limited number of studies have specifically explored ship detection in MR imagery. Most of these works tend to concentrate on specific aspects. For example, [14] delves into ship detection and characterization in favorable conditions. [15] addresses the issue of scarce annotated ship data and introduces a 'self-supervised learning' approach. Additionally, [16] introduces a method for identifying particular ship shapes associated with migrant activity. Few articles adopt a comprehensive approach [17], which is our primary area of interest.

This scarcity of studies is, in part, attributable to the absence of freely available ship datasets. In contrast to ship detection at high resolution, which benefits from an established reference dataset [18], the only publicly available Sentinel-2 ship datasets are [19] and [15]. These datasets comprise 31 (2000 ship exemplars) and 16 (1053 ship exemplars) Sentinel-2 images respectively, which may be limited to deploy state-of-the-art learning-based frameworks.

Only [17] addresses the detection of ships in medium-resolution satellite optical images within a relatively general framework. In this article, the analyzed images are sourced from both the Sentinel-2 mission satellites and the Planet Labs Dove satellite constellation. The images are divided into 800×800 pixels patches, which are then used as input for the Faster R-CNN [20] object detection model. The annotation process relies on colocating the satellite images with AIS (Automatic Identification System) data [21]. AIS data comprise ship identifiers and locations to groundtruth bounding boxes in the considered dataset. As most small ships are not equipped with AIS systems, this study only addresses large ships (> 100 m). Overall, this study reports a 85% detection rate but does not document the false alarm rate. These results do not appear fully conclusive compared with the performance reported for HR and SAR imagery.

The objective of our work is to create an automated system for detecting and characterizing ships in Sentinel-2 images.

III. Proposed approach

We present a multi-stage approach for ship detection and characterization in Sentinel-2 images. In the initial stage, we employ a sliding window technique to systematically cover the image. Each window is of size 100×100 pixels, representing 1 km $\times$ 1 km, and overlaps with neighboring windows by 25%. This overlap ensures that if a ship extends across two windows, a significant portion of it remains detectable in at least one of them. A window is considered valid if it contains at least 5% of sea pixels, determined using the land/sea mask. Subsequently, these valid windows are categorized as either "Ship" or "No ship" using a Resnet-type [22] classifier. For the windows classified as "Ship", we utilize a Faster R-CNN detector to obtain a bounding box around the detected ships (only the coordinates of the ships are necessary). This detector incorporates a dedicated branch to estimate the ship's heading. Figure 2 provides an overview of our ship detection and heading estimation system.

When different adjacent patches detect parts of the same ship, we apply the non-maximum suppression (NMS) algorithm. However, in our approach, the suppression is not based on network confidence but rather on the size of the bounding box. Our primary objective in this context is to retain the largest bounding box to obtain the most precise ship coordinates. Finally, once the coordinates are identified, we create a 50×50 image centered on these coordinates to estimate the ship length using a Resnet-type network (see Figure 3).

The motivation behind proposing a two-stage strategy is as follows. In current state-of-the-art schemes, candidate proposals are typically classified based on their internal features. Consequently, a region containing a very small vessel may exhibit the same features as a region containing a portion of rough sea. The context becomes crucial in this scenario, and the initial classification step is necessary to capture the context within sufficiently large patches that encompass the entire context while remaining small enough to avoid missing very small vessels.

In our approach, each individual component undergoes distinct training phases, following a traditional methodology that includes training, validation, and testing. It is strongly recommended to include a substantial number of complete Sentinel-2 images during the model development phase. This approach not only allows for the identification and resolution of potential methodological limitations but also facilitates a comprehensive examination of the test dataset's distribution.

A. Classification phase

Developing a "Ship" and "No ship" category classifier involves two main phases. Firstly, we focus on generating

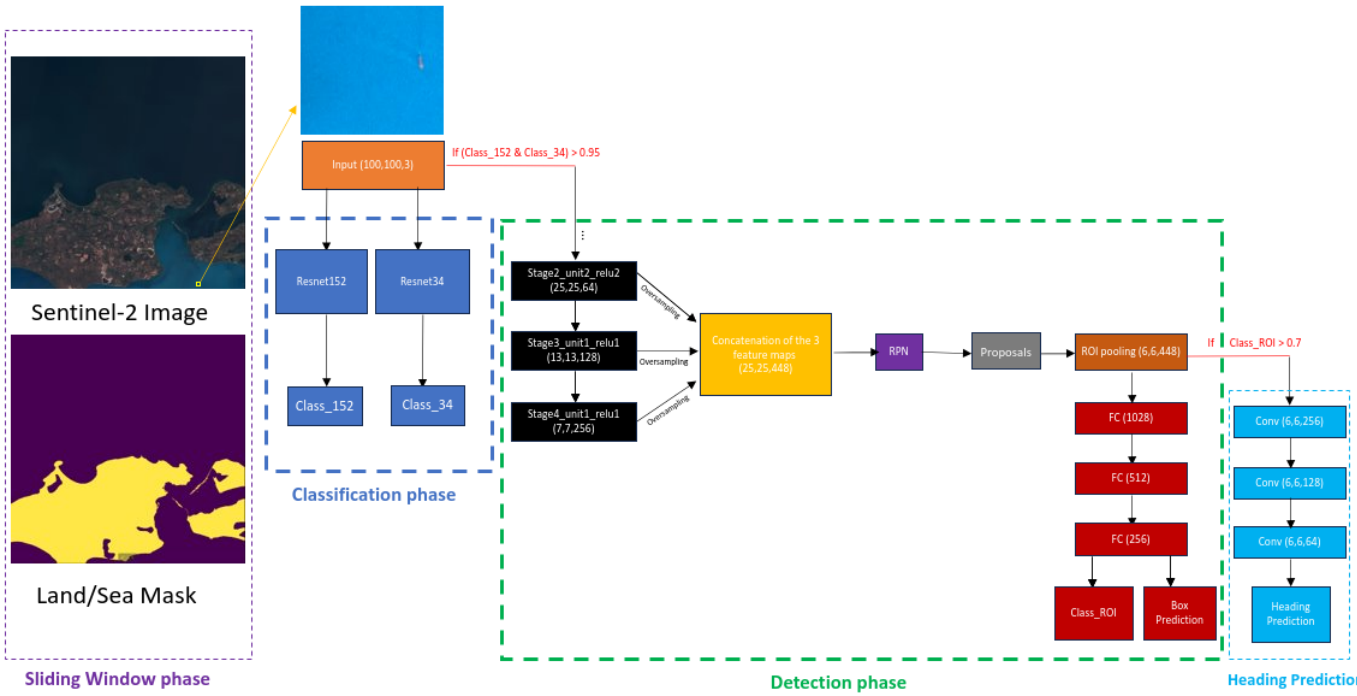


Fig. 2: An Overview of Our Ship Detection and Heading Estimation System on Sentinel-2 Images. The Faster R-CNN model comprises a backbone constructed from the Resnet18 classifier trained on Sentinel-2 data. The layer names in black used to build detection feature map are derived from keras applications [23].

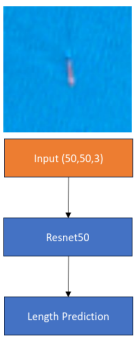


Fig. 3: Ship Length Estimation Component.

datasets and verifying their distribution. Secondly, we work on the model aspect, which includes selecting models and tuning parameters.

1) Considered datasets: We generate our dataset by extracting patches with dimensions of 100×100 pixels from Sentinel-2 images. Specifically, we utilize only the Red, Green, and Blue (RGB) bands (10 m resolution). This selection aims to leverage the advantages of a pre-trained neural network which is a powerful tool for increasing the accuracy and the robustness of our classifier [24]. These images are generated using the vessel detection reports provided by CLS analysts. These vessel detection reports are routinely generated by CLS in the context of its commercial activities and contains for each Sentinel-2 satellite image analysed the position of the detected vessels along its characteristics such as length and heading when these are measurable. Using the vessel detection reports available, we process over 6000 Sentinel 2 images, extracting approximately 12000 ship images and 20000 non-ship images. To our knowledge, this represents the first instance where a significantly large dataset has been employed for ship detection in medium-resolution optical imagery. The "No ship" category receives particular attention, as it must encompass all the scenarios encountered in optical images, including sea, land, and clouds. Images consisting solely of land are excluded, as their presence does not impact performance due to the availability of a land mask that eliminates these false alarms. Furthermore, it is essential to ensure that our dataset adequately represented objects resembling ships, such as small clouds, small islands, rough seas, and platforms. Manual verification is performed extensively to confirm the presence of all these "sub-categories" within our dataset.

2) Verification of Distribution: The partitioning of training, validation, and test sets may appear straightforward in the majority of machine learning problems. However, when dealing with automation applied to a wide range of diverse images under various scenarios, a careful partitioning becomes crucial. Let us assume that our network is trained on images featuring rough seas but is not tested on samples from this specific scenario to assess its "understanding". This can lead to vulnerabilities during the deployment phase: results on the test set may appear very promising, but performance on real-world images may be significantly poorer. Indeed, our issue involves binary classification. However, given the diversity of scenarios, we consider working with subcategories. In our work, We divide the "Ship" category into 11 sub-categories: Ship 10, Ship 20, Ship 30, Ship 40, Ship 50, Ship 100, Ship 150, Ship 200, Ship 300, Ship 400, and Ship 458. In this terminology, "Ship x" signifies a ship with a length less than "x". Figure

4 illustrates the distribution of "Ship" category in our dataset based on their size.

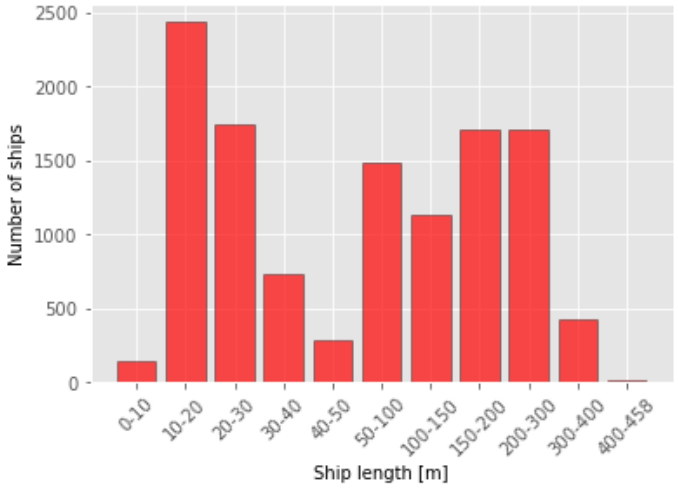


Fig. 4: Distribution of ship sub-categories.

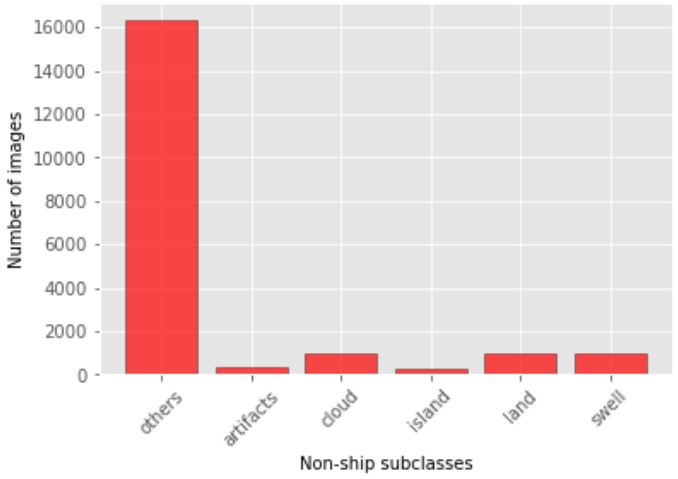
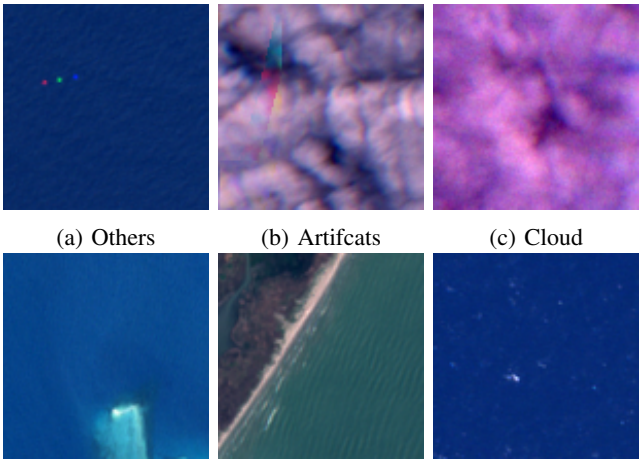


Fig. 5: Distribution of No ship sub-categories.

We also subdivide the "No ship" category into 6 sub-categories: **Others** compries patches devoid of vessels, randomly sampled from 6000 Sentinel-2 images, **Artifacts** contains patches with unwanted visual anomalies or distortions that can occur during image acquisition or processing, **Cloud** contains various cloud formations, **Island** includes patches featuring islands and underwater reliefs, **Land** predominantly consists of coastal areas, ports, and a few images with only land terrain, lastly, **Swell** contains various forms of swell. Figure 5 illustrates the distribution of "No ship" category based on this subdivision. Figure 6 shows some exemples of each sub-category.

To reduce the random variability caused by the train/test split, we further divide each of these subclasses into 5 clusters using an automatic classification method, specifically K-means [25]. This clustering is based on the histogram values of the image's 3 color channels. Each of these clusters is



(a) Others (b) Artifcats (c) Cloud

(d) Island (e) Land (f) Swell

Fig. 6: Examples of No ship sub-categories images.

subsequently divided into training, evaluation, and test sets. This approach ensures that the neural network is trained and tested on a wider range of data, resulting in a model that is not only more accurate but also more robust. Our approach is summarized in Diagram 7.

We also extensively employ data augmentation techniques to ensure a balanced representation of all our sub-cateogies. For the 'Ship' category, we apply vertical and horizontal flips as well as rotations. Regarding the 'No ship' category, we employ the same augmentation techniques in addition to elastic deformation to diversify the shapes of islands and clouds. It's worth noting that conducting a comprehensive enumeration of all the natural sub-categories is of paramount importance. The more comprehensive the enumeration, the more effective the network will be in practical applications. Carrying out this enumeration during the data generation phase can save considerable time and effort, as it reduces the need for frequent fine-tuning of the network.

3) Classification Network: We conduct extensive experiments using various pretrained models on the Imagenet dataset [26], including InceptionV3 [27], Resnet18, Resnet34, and Resnet152 [22], while fine-tuning different parameters. Our findings consistently favore the Resnet architecture, which demonstrates superior performance on our dataset.

In our training process, binary cross-entropy serves as our loss function. We utilize the Adam Optimizer with a batch size of 128 and a learning rate set at 10^{-4}.

To further enhance our model's ability to reduce false alarms, we implemented an ensemble approach using two classifiers: 'Resnet152' and 'Resnet34.' These two models exhibit comparable results and show significant improvements in false alarm reduction, albeit with a slight decrease in ship detection performance.

B. Detection phase

The core of our system is the classification component, which plays a pivotal role. The main goal of the detection component is to assign bounding boxes to ships that

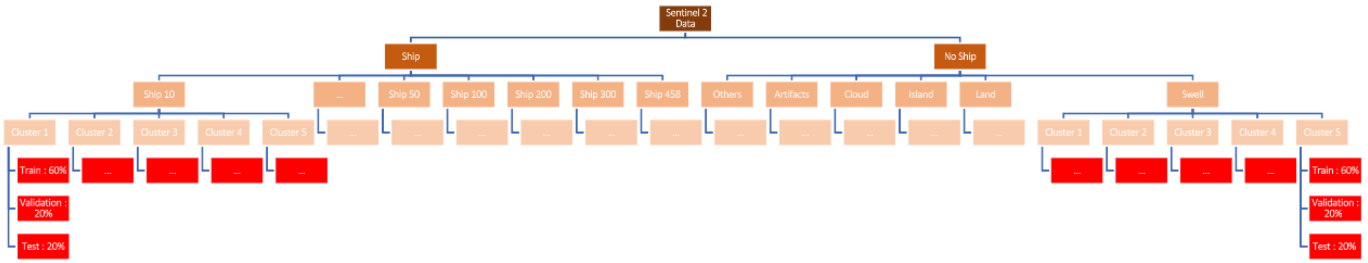


Fig. 7: Our K-means method for forming the training, validation, and test sets.

our classifier identifies within patches labeled as 'Ship'. Additionally, it may eliminate some false alarms if no ship is detected in the patch by the detector. To build our detection network, we employe the Faster R-CNN framework [20]. Despite its introduction in 2015, Faster R-CNN, along with adapted versions, continues to demonstrate its state-of-the-art performance on established reference datasets [28],[29]. However, it's important to note that we do not explore other detector types in this work, including YOLO and its variants [30] and detection transformer (DETR) and its variants [31].

We utilize a modified version of Faster R-CNN for detection, based on the code referenced in [32], implemented on the TensorFlow [33] framework. The primary motivation for these modifications is to enhance the detection of small vessels. The Faster R-CNN architecture can be divided into three key components: the Backbone network, the Region Proposal Network (RPN), and the Detector network. Since we are dealing with a single object type, we can use only the RPN along with the backbone network to detect ships. However, by employing the entire structure, we achieve significantly improved results. This improvement is attributed to the Detector network, which incorporates Fully Connected layers, refining the outcomes of the RPN, which relies solely on convolutional operations.

Our patches are sized at 100×100 pixels and contain sometimes very small vessels, representing just 1 pixel of the image, and sometimes large ships, occupying more than 40 pixels. To accommodate this size range, we configure the Region Proposal Network (RPN) to propose different regions of interest, even for the smallest objects. To achieve this, we utilize anchor boxes of various sizes based on the statistics of the ship length in our dataset, we choose 4×4, 7×7, 10×10, 14×14, 18×18, and 28×28 for square-shaped anchor boxes. Additionally, we maintain the original aspect ratios from the Faster R-CNN paper [20], which are 1, 0.5, and 2. Furthermore, the input feature map for the RPN is set to a size of 25×25 (with a stride length of 4) to preserve crucial information from very small vessels, especially in challenging scenarios. This configuration allows the RPN to effectively handle objects of different sizes within our patches.

The feature map is constructed from the classification network mentioned earlier. In fact, the classification network demonstrates excellent results in distinguishing between "Ship" and "No ship" images. As a result, we believe that the feature maps generated by this network encompass the essential information required to serve as the backbone of our detector. While the use of the initial 25×25 feature map from the classification network aids in proposing smaller regions that may contain small ships, it remains relatively shallow for capturing information about larger vessels and other objects within the image. To address this, we integrate intermediate feature maps of sizes 13×13 and 7×7. By employing upsampling techniques, we combine these feature maps to create the feature map for region proposals, ensuring a comprehensive representation of potential objects of interest [34].

The region proposal network (RPN) architecture, after creating the feature maps, remains unchanged from the original Faster R-CNN. It consists solely of a convolutional layer (512 channels) and two additional convolutional layers, one for classification and the other for regression.

During the training of the RPN, an anchor is considered positive only if its Intersection over Union (IOU) with a ground truth box exceeds 0.7. Conversely, it is labeled as negative if its IOU falls below 0.3. Anchors falling within this range are not used for training purposes. To address the class imbalance issue, we employ a weighted loss function, which helps balance the small number of positive anchors per image.

After proposing the regions of interest, we apply ROI pooling to transform these regions into a fixed size of 6×6, which is determined based on the lengths of the ships in our datasets. Subsequently, these resized regions are forwarded to the detector network. The structure of our modified Faster R-CNN is illustrated within the green rectangle in Figure 2.

Only the RPN part and the Detector part are trained with a learning rate of 10^{-4}. The weights of the backbone layers are frozen. We believe that since the classifier performs exceptionally well on a large dataset containing both "Ship" and "No ship" images, the network only needs to learn how to express the precise position of the ship, which is what the RPN and Detector are designed to accomplish.

C. *Characterization of ships*

In this section, we introduce the deep learning component for the estimation of the heading of the ship by adding an additional branch to the Faster R-CNN architecture. Additionally, we use a ResNet type model to estimate the size of the ship.

1) Headings: While it was feasible to train a bounding box with the ship's orientation, as demonstrated in [35], we have

access to a dataset containing 3000 ships annotated with their heading. This information proves to be more valuable than a simple rotated bounding box for our specific task.

Concerning the training process, we keep all the weights of our previously discussed Faster R-CNN model frozen. Then, we introduce an additional branch after the ROI pooling layer, as illustrated within the sky blue rectangle in Figure 2, to handle ship orientation estimation. In this branch, we utilize both the cosine and sine of the ship's angle as model outputs. This approach is adopted to convey circular information effectively to the network.

2) Length Estimation: Since regions of interest (ROIs) are resized during the process, some size information is inevitably lost. To address this issue, one approach is to recover size information using ship orientation and the scaling factor. However, given the current resolution, attempts to introduce an additional branch for length estimation, similar to what we do for ship heading, do not yield satisfactory results. Therefore, we decided to employ a separate network for this purpose.

We utilize the Resnet50 network, with inputs sized at 50×50 pixels and centered on the ship. While smaller patches can be considered (46 since the biggest ship we can meet is 458), during the deployment of the entire ship detection system, the inputs for the length estimation model are constructed around the center of the ship's bounding box predicted by our detector. This bounding box may be slightly affected by ship wakes or irregular ship shapes. Hence, we opt for this patch size, taking into account potential small errors in the detection phase.

It is essential to note that for ensuring the robustness of our approach, we systematically exclude all images featuring partial cropping and inadequately labeled lengths. However, we do not exclude images of ships partially obscured by clouds if their lengths are still measurable; otherwise, they are excluded. This refinement leads to a dataset consisting of 7460 images, distributed randomly across training (80%), validation (10%), and test (10%) sets.

IV. Experiments and Results

In this section, we present the outcomes of our system, along with experiments demonstrating the effectiveness of the methods used in both the classification and detection phases. We commence by introducing the experiments and results of the classification phase, followed by those of the detection phase. Subsequently, we delve into the characterization phase, and finally, we present the results of our complete ship detection system. To assess the performance of our system, we utilize a distinct dataset comprising 60 Sentinel-2 images, herein referred to as the **Evaluation set**. It is essential to note that this dataset is distinct from the training, validation, and test datasets used for training each component of the system

A. Description of the Evaluation Set

The evaluation set consists of 60 Sentinel-2 images captured from approximately 20 different Earth locations, as indicated by red stars in Figure 8, at various acquisition times. These images encompass a wide range of scenarios, including coastlines, turbulent seas, artifacts, images with a high density of ships, low-density ship images, and cloud-covered scenes. In total, the 60 images collectively contain 878475 valid patches. A patch is considered valid if it contains at least 5% of sea pixels. In total, the 60 images collectively contain around 1147 ships. The image annotations are carried out by CLS analysts. It's important to note that very small ships may not have been annotated due to the inherent difficulty and resolution limitations. Our evaluation set is available in [36].

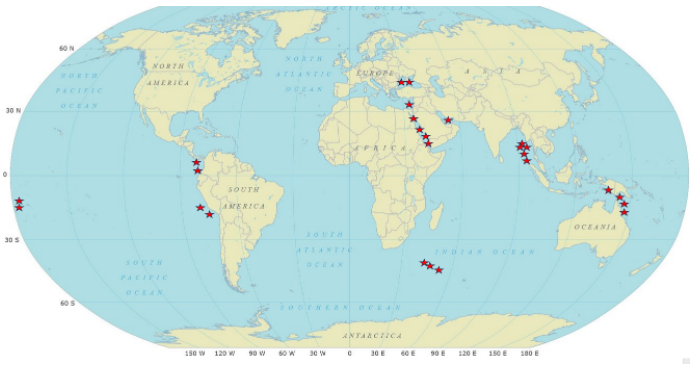


Fig. 8: Location of the 60 Sentinel 2 Images Constituting the Evaluation Set.

B. Classification Experimentations

1) Evaluating Distribution Verification Effectiveness: In this experiment, we analyze and compare the results achieved by employing various data splitting methods. We utilize the pre-trained Resnet152 network on the Imagenet dataset, as it exhibits the best performance. We examine three distinct data splitting configurations:

- In the first approach, we implement a random split, as depicted in the first phase of the diagram 7. This split is referred to as the "Random split."
- In the second approach, we employ a split based on subclasses, as shown in the second phase of the diagram 7. This split is referred to as the "Subclass split."
- In the third approach, we create a split based on clusters within each subclass, represented in the third phase of the diagram 7. This split is referred to as the "Subclass k-means split."

We present the performance metrics for the test sets in Table I, each comprising approximately 6500 images, with 2500 ship images and 4000 no-ship images. The metrics used to assess our classifiers include the False Positive Number (FP) and Recall. We opted for FP instead of precision due to the minimal occurrence of false positives. These metrics are compared across different confidence thresholds: 0.50, 0.75, and 0.95.

Table I shows comparative results; however, we cannot draw conclusions without testing the methods on a common dataset: the evaluation set (constructed using 60 Sentinel-2 images, as previously mentioned).

Table II displays the results of the three distinct models, evaluated on the **Evaluation set**. The FP metric represents the aggregate count of false positives observed across all 60 evaluation images, encompassing approximately 877470

Split / Threshold	0.5		0.75		0.95	
	FP	Rec.(%)	FP	Rec.(%)	FP	Rec.(%)
Random	45	98.1	34	97.7	15	96.0
Subclass	23	98.3	12	97.9	5	96.0
Subclass k-means	25	98.1	16	97.5	8	95.9

TABLE I: Evaluation of Data Splitting Methods on Their Respective Test Sets Using Resnet152.

Split / Threshold	0.5		0.75		0.95	
	FP	Rec.(%)	FP	Rec.(%)	FP	Rec.(%)
Random	2529	98.2	1818	97.6	1003	96.1
Subclass	1050	**98.4**	840	97.9	605	96.2
Subclass k-means	546	98.2	406	97.1	**240**	95.5

TABLE II: Evaluation of Data Splitting Methods on The Evaluation Set Using Resnet152.

valid patches devoid of ships. Remarkably, the 'k-means split' method outperforms both the 'random split' and 'subclass split' approaches, resulting in the lowest occurrences of false positives, averaging approximately 4 per Sentinel-2 image. This outcome underscores the significance of employing verification techniques and methodological refinements to reduce the random variability inherent in splitting datasets, particularly in complex scenarios with a wide range of image characteristics.

2) Ensemble network classification: Throughout the remainder of our results, we exclusively employ the dataset split derived from the k-means method. To minimize the False Positive Number, we conduct extensive experiments with various models pretrained on Imagenet. Table III summarizes the outcomes of these models in terms of False Positive Number (FP) and Recall metrics.

Model / Threshold	0.5		0.75		0.95	
	FP	Rec.(%)	FP	Rec.(%)	FP	Rec.(%)
InceptionV3	36	97.6	25	97.0	11	95.0
Resnet18	40	97.0	28	96.9	13	94.0
Resnet34	33	97.5	24	97.2	9	94.8
Resnet152	25	**98.1**	16	97.5	8	95.9
Ens. Resnet34/Resnet152	14	96.5	10	95.3	**4**	93.0

TABLE III: Evaluation of Various Pretrained Models' Performance on the Test Dataset.

Our aim is to create the optimal ensemble, prioritizing FP reduction while tolerating a slight drop in Recall. To achieve this, we employ ensemble models and manually selected aggregation rules, guided by a criterion: False Positives should not include any images clearly devoid of vessels. Consequently, we choose Resnet152 and Resnet54 with an aggregation rule stipulating that both models must output a confidence score exceeding **0.95** to classify an image as containing a ship.

The Ensemble network yields a Recall of **93%** with only **4** false positives on the test set. Figure 9 illustrates images of these four false positives. It is not evident whether these images contain a ship.

Fig. 9: The only four false positives produced by the Resnet152 and Resnet34 ensemble.

C. Detection Experimentations

1) Evaluating our Modified Faster R-CNN: In this experiment, we assess the effectiveness of our modified Faster R-CNN in comparison to the original version. To ensure a robust comparison, we adjust the classical Faster R-CNN with a ResNet18 backbone pretrained on the ImageNet dataset, solely due to differences in input size. The feature map used for generating proposals is derived from the layer "Stage_4_unit1_relu1". To evaluate the performance of these models, we employ Detection rate, False alarm rate , F1 score, and Average Precision, noting that a detection is considered correct only if it's Intersection over Union (IOU) exceeds 0.5. Additionally, we conduct a performance comparison using ResNet152 and ResNet34 backbones. Our adapted version of the Faster R-CNN is denoted as MRS-Faster R-CNN, underscoring its focus on medium-resolution ship detection.

Model / Metric	D(%)	FA(%)	F1 score(%)	AP(%)
Faster R-CNN / ResNet18	70	66	45.7	30.9
MRS-Faster R-CNN / ResNet18	**80.4**	**18.8**	**75.4**	**77.1**
MRS-Faster R-CNN / ResNet34	77.5	23.7	76.9	74.4
MRS-Faster R-CNN / ResNet152	78	27.8	75	73.1

TABLE IV: Performance Metrics for Different detection Models.

Our MRS-Faster R-CNN demonstrates promising performance in comparison to a standard version of Faster R-CNN. Even though Resnet152 and Resnet34 yield superior classification results (Table III), the feature maps constructed from Resnet18 produce the best outcomes. We hypothesize that the comparatively shallower architecture of Resnet18 allows the intermediate feature maps which are of a relatively large size to encompass a richer set of information. It's worth noting that utilizing a backbone constructed from Resnet152 or Resnet34 for detection will help minimize image processing time in our system, albeit with a minor trade-off in performance.

Detection and False Alarm rates are influenced by two key factors. Firstly, our dataset comprises numerous small vessels with wakes, leading to instances where wakes are considered as part of the ship due to resolution limitations. However, our detector consistently distinguishes between wakes and vessels on larger ships, as depicted in Figure 10. Secondly, our model encounters challenges in detecting closely positioned ships, particularly small vessels, as illustrated in Figure 11. Refining the Intersection over Union (IOU) condition to 0.25, Detection rate reaches up to **95%**, and False alarm rate decreases to **5%**.

2) Direct application of the Faster R-CNN: In this experiment, we directly apply our detector to the evaluation set,

Fig. 10: Model detections exemples : Green bounding boxes denote annotations, while yellow indicates model predictions.

Fig. 11: Model struggles with close ship differentiation, generating a single bounding box for closely positioned vessels.

excluding the use of the classification component, following the approch of [17]. This direct application yields a Detection Rate of **96%** and an Average False Alarm number per image of **50**. The expected outcome aligns with the inherent design of the Faster R-CNN, which lacks contextual information. To ensure a fair experiment with direct applications, additional enhancements are essential. Incorporating branches to encode contextual information, as demonstrated in [37], represents a potential avenue for improvement. Notably, such enhancements were not explored in our work.

D. System Performance evaluation

We assess the performance of our ship detection system, emphasizing the detection rate and false alarm count. Figure 12 presents the detection rates for various ship sizes, showcasing an overall detection rate of **93.2%**. The system records an average of **2.1** to **3.9** false alarms per Sentinel-2 image. The uncertainty in false alarm numbers stems from the presence of 106 detections resembling small ships ($<$ 20m) but not annotated by CLS analysts due to resolution limitations. Notably, our system excels in detecting large vessels ($>$ 100m), achieving a detection rate exceeding **97%**. Table V presents detailed results for each image, wherein false alarms are predominantly attributed to harbor facilities, offshore platforms, and swell (Figure 13). These results demonstrate significant promise for achieving complete automation in ship detection on Sentinel-2 images and, more broadly, on optical medium-resolution images.

E. Ship Characterization Results

1) Heading Results: Figure 14 displays the heading results for 600 ship images. We assess accuracy in two ways. In the left part of the figure, a predicted heading is considered accurate if the absolute difference between the predicted heading (h_p) and the ground truth heading (h_{gt}) is less than a specified heading error (h_{error}), with h_{error} taking values of 15°, 30°, and 45°. In the right part of the figure, accuracy is determined by verifying if the absolute difference between the predicted heading and the actual heading modulo 180° is less than h_{error}. This relaxation of the first condition helps accurately assess the model's capability, especially in situations where the direction is unclear due to resolution limitations, as depicted in Figure 15. The accuracy for predicting headings within a range of ± 30 degrees is 93%.

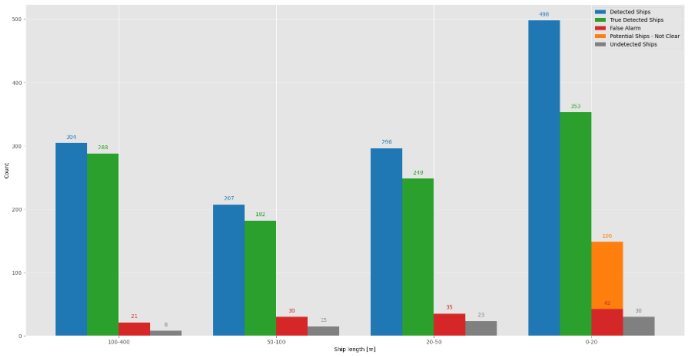


Fig. 12: System Performance evaluation : Overall Detection rate : 93.2%, Potential False Alarms (Range): 128-234, Average False Alarms for One Image (Range): 2.1-3.9.

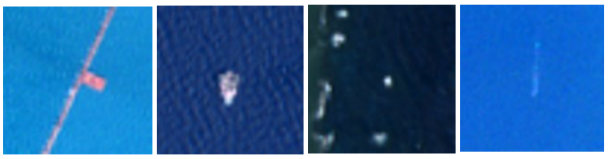


Fig. 13: Examples of False Alarms.

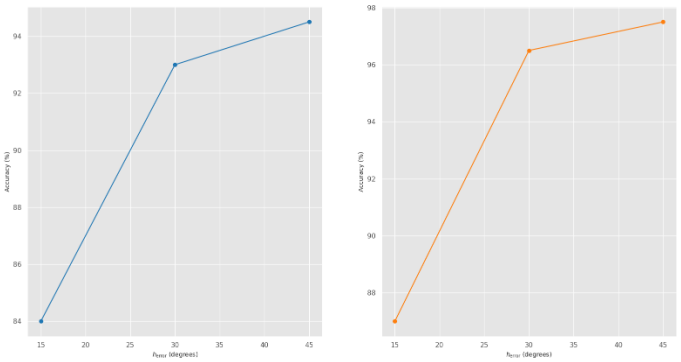


Fig. 14: Heading Prediction Results for 600 Ship Images. Left: A correct prediction is indicated when $|h_p - h_{gt}| \leq h_{\text{error}}$. Right: A correct prediction occurs when $|h_p - h_{gt}| \mod 180 \leq h_{\text{error}}$.

2) Length estimation Results: The results on length estimation are based on 746 ship images with dimensions of 50 $\times$ 50. The mean length error is reported as 15.36m $\pm$ 19.57m, accompanied by an R-squared value of 92 (Figure 16). The relatively high standard deviation error is attributed

Image Name	Main Features	Total Ships	DR(%)	FA	Potential Ships, Not Clear
S2A_MUL_ORT_13_20220521T084745_20220521T084745_00064	Coast, Calm sea	15	86.6	1	0
S2A_MUL_ORT_13_20220521T084756_20220521T084756_00064	Coast, Calm sea	13	100	0	7
S2A_MUL_ORT_13_20220702T063238_20220702T063238_00091	Clouds	0	.	0	0
S2A_MUL_ORT_13_20220707T083745_20220707T083745_00021	Coast, Calm sea, Harbor	140	94.2	6	3
S2A_MUL_ORT_13_20220712T063233_20220712T063233_00091	Clouds	0	.	0	0
S2A_MUL_ORT_13_20220719T074600_20220719T074600_00049	Coast, Islands	18	100	0	13
S2A_MUL_ORT_13_20220719T074613_20220719T074613_00049	Coast, Islands, Clouds	26	100	4	20
S2A_MUL_ORT_13_20220719T074628_20220719T074628_00049	Coast, Islands, Clouds, Underwater relief	15	93.3	4	3
S2A_MUL_ORT_13_20220719T074638_20220719T074638_00049	Coast, Islands, Clouds, Underwater relief, Artifacts	8	100	4	3
S2A_MUL_ORT_13_20220801T063228_20220801T063228_00091	Small Clouds	0	.	0	0
S2A_MUL_ORT_13_20220801T063229_20220801T063229_00091	Small Clouds, Islands	0	.	0	0
S2A_MUL_ORT_13_20220801T063235_20220801T063235_00091	Small Clouds, Islands	0	.	0	0
S2A_MUL_ORT_13_20220808T062238_20220808T062238_00048	Clouds	0	.	0	0
S2A_MUL_ORT_13_20220818T062238_20220818T062238_00048	Clouds	1	0	0	0
S2A_MUL_ORT_13_20220821T063236_20220821T063236_00091	Clouds	0	.	0	0
S2A_MUL_ORT_13_20220828T062238_20220828T062238_00048	Complete Cloud Cover	0	.	0	0
S2A_MUL_ORT_13_20220908T004106_20220908T004106_00059	Coast, Small Clouds, Underwater relief	14	85.7	0	2
S2A_MUL_ORT_13_20220908T004119_20220908T004119_00059	Coast, Small Clouds, Underwater relief	0	.	0	1
S2A_MUL_ORT_13_20220908T004130_20220908T004130_00059	Coast, Small Clouds, Underwater relief	16	81.3	5	0
S2A_MUL_ORT_13_20220910T063231_20220910T063231_00091	Calm Sea, Small Clouds	0	.	0	0
S2A_MUL_ORT_13_20220910T063234_20220910T063234_00091	Clouds	0	.	0	0
S2A_MUL_ORT_13_20220912T220946_20220912T220946_00129	Clouds, Swell	0	.	7	0
S2A_MUL_ORT_13_20220915T034601_20220915T034601_00018	Clouds, Platforms	14	100	2	5
S2A_MUL_ORT_13_20220915T034604_20220915T034604_00018	Clouds, Platforms	15	100	1	5
S2A_MUL_ORT_13_20220915T034630_20220915T034630_00018	Clouds	3	100	0	0
S2A_MUL_ORT_13_20220915T222001_20220915T222001_00029	Clouds	0	.	0	0
S2A_MUL_ORT_13_20220918T035449_20220918T035449_00061	Clouds	40	90	6	3
S2A_MUL_ORT_13_20220918T035521_20220918T035521_00061	Coast, Clouds	8	75	0	0
S2A_MUL_ORT_13_20220918T035607_20220918T035607_00061	Coast, Clouds, Islands	20	80	4	2
S2A_MUL_ORT_13_20220918T035616_20220918T035616_00061	Coast, Clouds	17	70	2	0
S2A_MUL_ORT_13_20220918T035622_20220918T035622_00061	Coast, Clouds, Islands	97	92.7	1	4
S2A_MUL_ORT_13_20220920T063234_20220920T063234_00091	Clouds	0	.	0	0
S2B_MUL_ORT_13_20220509T085722_20220509T085722_00107	Coast, Calm Sea	31	100	4	3
S2B_MUL_ORT_13_20220704T062230_20220704T062230_00048	Clouds	0	.	0	0
S2B_MUL_ORT_13_20220709T083234_20220709T083234_00121	Coast, Clouds, Underwater relief	69	92.7	6	1
S2B_MUL_ORT_13_20220709T083248_20220709T083248_00121	Coast, Calm Sea	13	100	1	1
S2B_MUL_ORT_13_20220710T145840_20220710T145840_00139	Coast, Clouds, Artifacts	8	100	2	2
S2B_MUL_ORT_13_20220717T063228_20220717T063228_00091	Clouds	0	.	0	0
S2B_MUL_ORT_13_20220718T155111_20220718T155111_00111	Coast, Clouds	0	.	2	0
S2B_MUL_ORT_13_20220719T083234_20220719T083234_00121	Coast, Calm Sea, Underwater relief	87	91.9	1	2
S2B_MUL_ORT_13_20220719T083238_20220719T083238_00121	Calm Sea, Underwater relief, Platforms	147	91.8	10	11
S2B_MUL_ORT_13_20220722T070213_20220722T070213_00020	Coast, Clouds, Swell	129	96.8	34	5
S2B_MUL_ORT_13_20220727T063228_20220727T063228_00091	Clouds	0	.	0	0
S2B_MUL_ORT_13_20220813T062230_20220813T062230_00048	Clouds	0	.	0	0
S2B_MUL_ORT_13_20220816T063227_20220816T063227_00091	Clouds	1	0	0	0
S2B_MUL_ORT_13_20220828T083001_20220828T083001_00121	Coast, Harbor, Calm Sea	107	97.1	9	3
S2B_MUL_ORT_13_20220828T083030_20220828T083030_00121	Coast, Clouds	31	74.1	3	0
S2B_MUL_ORT_13_20220909T005948_20220909T005948_00002	Small Clouds	0	.	0	0
S2B_MUL_ORT_13_20220909T005951_20220909T005951_00002	Small Clouds	1	100	1	0
S2B_MUL_ORT_13_20220909T005956_20220909T005956_00002	Clouds, Islands	7	57	2	0
S2B_MUL_ORT_13_20220909T005959_20220909T005959_00002	Calm Sea, Clouds	5	100	0	0
S2B_MUL_ORT_13_20220909T010014_20220909T010014_00002	Calm Sea, Clouds	5	80	0	0
S2B_MUL_ORT_13_20220909T010028_20220909T010028_00002	Coast, Calm Sea	1	100	0	2
S2B_MUL_ORT_13_20220912T062227_20220912T062227_00048	Clouds	0	.	0	0
S2B_MUL_ORT_13_20220916T040541_20220916T040541_00104	Coast, Clouds	4	50	2	1
S2B_MUL_ORT_13_20220916T040545_20220916T040545_00104	Clouds, Islands	0	.	0	2
S2B_MUL_ORT_13_20220916T040610_20220916T040610_00104	Coast, Clouds	11	90	2	1
S2B_MUL_ORT_13_20220916T040624_20220916T040624_00104	Clouds	3	100	1	0
S2B_MUL_ORT_13_20220916T040628_20220916T040628_00104	Clouds	1	100	1	1
S2B_MUL_ORT_13_20220916T040634_20220916T040634_00104	Clouds	6	100	0	0

TABLE V: Detection System's performance on the Evaluation Set [36] with Detection Rate and False Alarm Count metrics.

Fig. 15: Examples of Ambiguous Heading.

to instances where wake shapes or partial cloud coverage make the length measurement task challenging (Figure 17). By excluding these challenging instances, the dataset now consists of 695 images, yielding a reduced mean length error of 11.29m $\pm$ 10.17m. Despite these challenges, our results demonstrate commendable performance, particularly considering the resolution limitations of Sentinel-2 data.

V. Discussion

The detection of ships in medium resolution optical images remains an underexplored field, with only a few studies adopting a comprehensive approach to this task. The scarcity of research in this area is not indicative of its insignificance but rather stems from the limited availability of data and annotated resources. In an attempt to bridge this gap, we

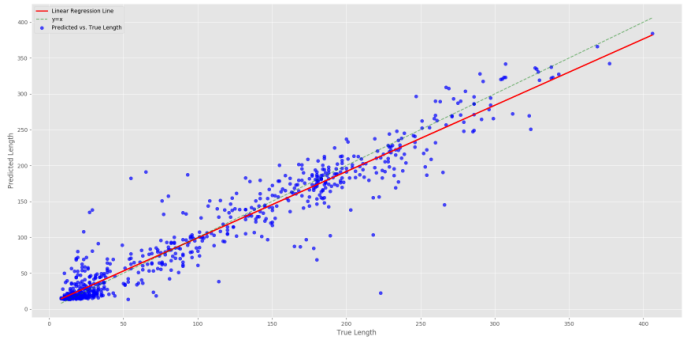
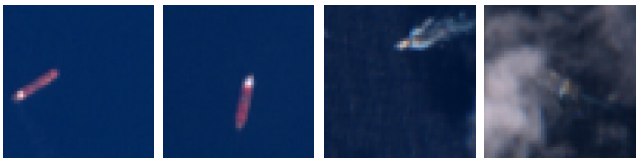


Fig. 16: Predicted vs. True ship lengths with an R-squared of 0.92.



(a) 145(142) (b) 177(173) (c) 68(180) (d) 218(103)

Fig. 17: Examples of Ship Length Prediction. The first number denotes the ground truth length, while the predicted length is enclosed in parentheses

make available 60 annotated Sentinel-2 images (1147 ship exemplars) which include length, and heading information when theses are mesurable [36].

Our study demonstrates a significant improvement in performance compared to existing results: our detection rate is 93%, compared to 75% in [15]. Notably, our detection rate on large vessels ($>$ 100 m) is 97% compared to 85% in [17]. It is crucial to highlight that this improvement is primarily associated with the substantial difference in the dataset sizes rather than the methodology employed. Importantly, it should be noted that reproducing our results does not necessitate an equivalent volume of data. Our dataset contains a substantial amount of redundant information, suggesting that achieving comparable results may be feasible with a more carefully selected dataset.

It's crucial to highlight that our primary research objective centers around establishing an effective ship detection system, with ship heading and length estimation emerging as subsequent considerations. While there is room for improvement, particularly in length estimation (our mean error: 15.36m $\pm$ 19.57m on 746 ship images), in [14], the reported mean error is 15.48m $\pm$ 10.94m on 34 ships, only 3 of which contain ship wakes. Our higher standard deviation error arises from the presence of ships partially obscured under clouds and a significant number of images with ship wakes. Upon removing these challenging images, the mean length error is reduced to 11.29m $\pm$ 10.17m.

The results for heading estimation have proven highly satisfactory. We report a 93% accuracy in heading estimation, where correct heading estimation is defined as ground truth heading $\pm$30 degrees. The remaining 7% discrepancy is attributed to instances where the heading is visually unclear due to resolution limitations. It is noteworthy that, to the best of our knowledge, we did not find any studies presenting results in heading estimation specifically on medium-resolution (MR) optical imagery.

The demonstrated results hold promise for achieving complete task automation. However, it is essential to note that full confirmation of these findings awaits deployment in production settings (Figure 18), considering the diversity and occasional complexity of scenes encountered. Our system exhibits a time limitation, requiring approximately 10 minutes to process a complete Sentinel-2 image on an 8-Core CPU. This processing time is attributed to the multiple stages involved, which, with the application of alternative ideas, could potentially be consolidated into a single stage without compromising performance.

One potential avenue for system enhancement involves introducing a context branch to the Faster R-CNN architecture [37]. This modification holds promise in addressing issues such as false alarms triggered by large vessels, which may generate noise in their surroundings, producing bright points that our detector may misinterpret as small ships. To overcome the time limitations inherent in our system, the utilization of a one-stage detector, such as YOLO [38], could be considered. Additionally, we acknowledge the challenge of accurately detecting very close ships, reflecting a broader issue in general object detection [39]. Future research efforts may explore innovative strategies to mitigate these challenges and further refine the performance of ship detection in optical images.

VI. Conclusion

In this research, we introduced an automated system designed to detect and characterize ships in Sentinel-2 images through a multi-stage process. Our system comprises distinct components focusing on classification, detection, and characterization. Within the classification component, we highlighted the importance of validating distribution in complex environments and diverse scenarios. State-of-the-art models, specifically Resnet152 and Resnet34, were employed to minimize false positives. During the detection phase, modifications were made to the original Faster R-CNN to effectively handle our specific dataset. In the characterization phase, a dedicated branch was integrated into the Faster R-CNN to estimate ship headings, and Resnet50 was separately used for length estimation. The results of our system are promising, indicating the potential for complete automation of ship detection and characterization in Sentinel-2 images. Additionally, the presented approach is adaptable to other satellite constellations. However, our system encounters challenges in differentiating small ships from their wakes and struggles with detecting closely positioned ships. Acknowledging these limitations, further research could explore these specific points, ensuring a continuous improvement in automated ship detection and characterization methodologies.

VII. Acknowledgements

The dataset employed in this paper was provided by Collecte Localisation Satellites (CLS). Our sincere appreciation goes to

Fig. 18: Integration of the Ship Detection System in CLS Processing Pipeline - Initial System Trials.

Vincent Kerbaol for his pivotal support, making this project possible. We express our gratitude to the CLS analysts for their meticulous annotation of the datasets.

Special recognition is extended to Etienne Gauchet, Mohammed Benayad, and Maxime Vandevoorde for their contributions to data cleaning and generation. We offer a heartfelt thanks to Chaïmae Sriti and Aurélien Colin for their comprehensive review of the paper and the insightful comments they provided.

REFERENCES

[1] EMSA, "Emsa outlook 2023," 2022.

[2] A. Iodice and G. Di Martino, *Maritime Surveillance with Synthetic Aperture Radar*. Institution of Engineering and Technology, 2020.

[3] V. W. G. H, *Satellite Imaging for Maritime Surveillance of the European Seas*. Berlin (Germany): Springer Science+Business Media B.V., 2008.

[4] D. J. Crisp, "The state-of-the-art in ship detection in synthetic aperture radar imagery," tech. rep., Defence Science and Technology Organisation Salisbury (Australia) Info ..., 2004.

[5] L. Zhai, Y. Li, and Y. Su, "Segmentation-based ship detection in harbor for sar images," in *2016 CIE International Conference on Radar (RADAR)*, pp. 1–4, 2016.

[6] M. Stasolla, J. J. Mallorqui, G. Margarit, C. Santamaria, and N. Walker, "A comparative study of operational vessel detectors for maritime surveillance using satellite-borne synthetic aperture radar," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 6, pp. 2687–2701, 2016.

[7] B. LI, X. XIE, X. WEI, and W. TANG, "Ship detection and classification from optical remote sensing images: A survey," *Chinese Journal of Aeronautics*, vol. 34, no. 3, pp. 145–163, 2021.

[8] S. Voinov, D. Krause, and E. Schwarz, "Towards automated vessel detection and type recognition from vhr optical satellite images," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 4823–4826, 2018.

[9] W. Chen, B. Han, Z. Yang, and X. Gao, "Mssdet: Multi-scale ship-detection framework in optical remote-sensing images and new benchmark," *Remote Sensing*, vol. 14, no. 21, 2022.

[10] U. Kanjir, H. Greidanus, and K. Oštir, "Vessel detection and classification from spaceborne optical images: A literature survey," *Remote Sensing of Environment*, vol. 207, pp. 1–26, 2018.

[11] S. Zhang, R. Wu, K. Xu, J. Wang, and W. Sun, "R-cnn-based ship detection from high resolution remote sensing imagery," *Remote Sensing*, vol. 11, no. 6, 2019.

[12] G. Tang, S. Liu, I. Fujino, C. Claramunt, Y. Wang, and S. Men, "H-yolo: A single-shot ship detection approach based on region of interest preselected network," *Remote Sensing*, vol. 12, no. 24, 2020.

[13] S. Karki and S. Kulkarni, "Ship detection and segmentation using unet," in *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, pp. 1–7, 2021.

[14] H. Heiselberg, "A direct and fast methodology for ship recognition in sentinel-2 multispectral imagery," *Remote Sensing*, vol. 8, no. 12, 2016.

[15] A. Ciocarlan and A. Stoian, "Ship detection in sentinel 2 multi-spectral images with self-supervised learning," *Remote Sensing*, vol. 13, no. 21, 2021.

[16] U. Kanjir, "Detecting migrant vessels in the mediterranean sea: Using sentinel-2 images to aid humanitarian actions," *Acta Astronautica*, vol. 155, pp. 45–50, 2019.

[17] D. Štepec, T. Martinčič, and D. Skočaj, "Automated system for ship detection from medium resolution satellite optical imagery," in *OCEANS 2019 MTS/IEEE SEATTLE*, pp. 1–10, 2019.

[18] S. Kızılkaya, U. Alganci, and E. Sertel, "Vhrships: An extensive benchmark dataset for scalable deep learning-based ship detection applications," *ISPRS International Journal of Geo-Information*, vol. 11, no. 8, 2022.

[19] F. D. Vieilleville, A. Lagrange, N. Dublé, and B. L. Saux, "Sentinel-2 dataset for ship detection," Mar. 2022.

[20] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks.," in *NIPS* (C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, eds.), pp. 91–99, 2015.

[21] A. Milios, K. Bereta, K. Chatzikokolakis, D. Zissis, and S. Matwin, "Automatic fusion of satellite imagery and ais data for vessel detection," in *2019 22th International Conference on Information Fusion (FUSION)*, pp. 1–5, 2019.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[23] F. Chollet *et al.*, "Keras applications." https://keras.io/api/applications/, 2015.

[24] D. Hendrycks, K. Lee, and M. Mazeika, "Using pre-training can improve model robustness and uncertainty," *CoRR*, vol. abs/1901.09960, 2019.

[25] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, University of California Press, Berkeley*, pp. 281–297, 1967.

[26] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[27] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, vol. abs/1512.00567, 2015.

[28] K. He, R. B. Girshick, and P. Dollár, "Rethinking imagenet pre-training," *CoRR*, vol. abs/1811.08883, 2018.

[29] T. Mahendrakar, A. Ekblad, N. Fischer, R. White, M. Wilde, B. Kish, and I. Silver, "Performance study of yolov5 and faster r-cnn for autonomous navigation around non-cooperative targets," in *2022 IEEE Aerospace Conference (AERO)*, pp. 1–12, 2022.

[30] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.

[31] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I*, (Berlin, Heidelberg), p. 213–229, Springer-Verlag, 2020.

[32] B. Trzynadlowski, "Faster r-cnn." https://github.com/trzy/FasterRCNN, 2022.

[33] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[34] C. Cao, B. Wang, W. Zhang, X. Zeng, X. Yan, Z. Feng, Y. Liu, and Z. Wu, "An improved faster r-cnn for small object detection," *IEEE Access*, vol. 7, pp. 106838–106846, 2019.

[35] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, "Oriented r-cnn for object detection," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3500–3509, 2021.

[36] M. Bou-laouz, R. Vadaine, G. Hajduch, and R. Fablet, "Sentinel-2 dataset for ship detection and characterization," Dec. 2023. https://doi.org/10.5281/zenodo.10222276.

[37] J. Leng, Y. Liu, T. Zhang, and P. Quan, "Context learning network for object detection," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 667–673, 2018.

[38] J. Terven and D. Cordova-Esparza, "A Comprehensive Review of YOLO: From YOLOv1 and Beyond," *arXiv e-prints*, p. arXiv:2304.00501, Apr. 2023.

[39] R. Kaur and S. Singh, "A comprehensive review of object detection with deep learning," *Digital Signal Processing*, vol. 132, p. 103812, 2023.