ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

Ship detection

Machine Learning and Content Analytics

Bellos Adam (f2822009)
Nakos Antonios (f2822010)
Ziou Alexandros (f2822020)
September 2021

Contents

Introduction

Maritime Domain Awareness is the effective understanding of anything associated with the maritime domain that could impact the security, safety, economy, or environment. With the current state of global affairs, worldwide tensions are high, creating a landscape in which commercial companies, financial institutions, governments and flag administrations must carefully monitor the security of their supply chains and maritime domains for sanctions and environmental regulation violations, the use of deceptive shipping practices, and suspicious, criminal, or terrorist activity.

In this context, Maritime Surveillance is regarded as the most essential practice for creating Maritime Awareness. Today's highly complex economic and political landscape has rendered Maritime Surveillance a significant challenge, especially in regions where maritime traffic represents major economic interests bound to potential threats and illegal activities such as smuggling, illegal immigration or maritime piracy. It goes without saying that traditional surveillance means such as patrol vessels and aircrafts are not sufficient for efficient Maritime Surveillance, since they constitute a relatively scarce and costly recourse, especially for large-scale areas with various activities to be monitored. Therefore, it has become imperative for law enforcement, border and security authorities, fisheries management agencies and other organizations with interests in the maritime domain, to take advantage of state-of-the-art technologies with the aim to enhance Maritime Surveillance in the most effective and cost-efficient way possible.

Our project, vision and goals

For our project, we focused on one of the most crucial aspects of Maritime Surveillance in today's highly technologically advanced landscape, which is Ship Detection on remote sensing images, using a deep learning approach. The main functions of a Ship Detection system include monitoring traffic flow, detect suspicious activity such as pirate threats, illegal fisheries and smuggling, and complement the prevention of marine pollution. Thus, such a system can play a vital role in maritime security management and even national security, providing essential information for maritime awareness and strategic decision-making in a wide range of operational environments.

The development of remote sensing imaging technology combined with continuous innovations in the field of deep learning, particularly in objection detection algorithms, has made possible the development of models which can be deployed with the aim of monitoring wide-range sea surfaces, enhancing Maritime Surveillance. For our project, we used images that were generated using Synthetic Aperture Radar (SAR) technology. SAR is known for its high resolution over optical satellite images as well as its all-weather working capabilities, rendering it a valuable source of data for our project. Regarding our approach in building our Ship Detection model, we opted for a one-stage object detection method, basing our model on the Single Shot Detection (SSD) model, published by Wei Liu et al. in 2015. To make our decision, we took into consideration that for a Ship Detection model it is crucial to detect not only accurately enough but also as early as possible, since in many cases security is at stake as in the case of a pirate threat for instance. Thus, our main concern was to find the optimal trade-off between speed and accuracy. A model similar to SSD seemed to correspond to our standards as it is computationally efficient without sacrificing much accuracy, as demonstrated in the Wei Liu et al. paper.

Data Collection and Dataset Overview

The dataset that ushered in the project implementation phase consists of images of different types of ships and marine areas. After having been captured from satellites, the images were subjected to a fragmentation process, which led to the creation of the final form of the dataset. More specifically, the imagery was obtained using the most advanced commercially available earth observation radar image providers—the Synthetic Aperture Radar (SAR) images were taken by Sentinel-1 (extra wide swath) of the European space agency and by Radasat-2 of the Canadian space agency. The spaceborne photography was taken between 22 July and 6 October of 2015 and depicts approximately 80% of the South African exclusive economic area (EEZ), including coastal areas and ports where high ship density is observed. Finally, it should be noted that the pictures have sizes of 512x512 pixels.

Data Processing/Annotation/Normalization

The SAR images underwent processing that resulted in 1,596 sub-images containing ships and 500,000 not containing. As part of the data processing, a random sample of 3.192 sub-images was taken from the latter sub-set of the data (500,000 pictures). As a result, a more balanced dataset could be constructed, ensuring in this way the existence of substantial learning material for our model. If this intervention had not been made, the model would learn better the pattern for the negative class and end up classifying everything as not having a ship.

Afterward, the 1596 sub-images were annotated manually with an open-source tool called YOLO mark. This software performs image marking for machine learning and is an integral component of many object detection tasks. As shown in the following picture, the marking entails the placement of the object inside a rectangular frame and the assignment of a label, which in our case is 'ship'. The txt files that were created with this process were then converted to Pascal VOC file format. The Pascal Visual Object Classes (VOC) challenge is a benchmark in visual object category recognition and detection, providing the vision and machine learning communities with a standard dataset of images and annotation, and standard evaluation procedures. At last, we have a dataset with 4,788 observations out of which 1,596 or 1/3 include ships.

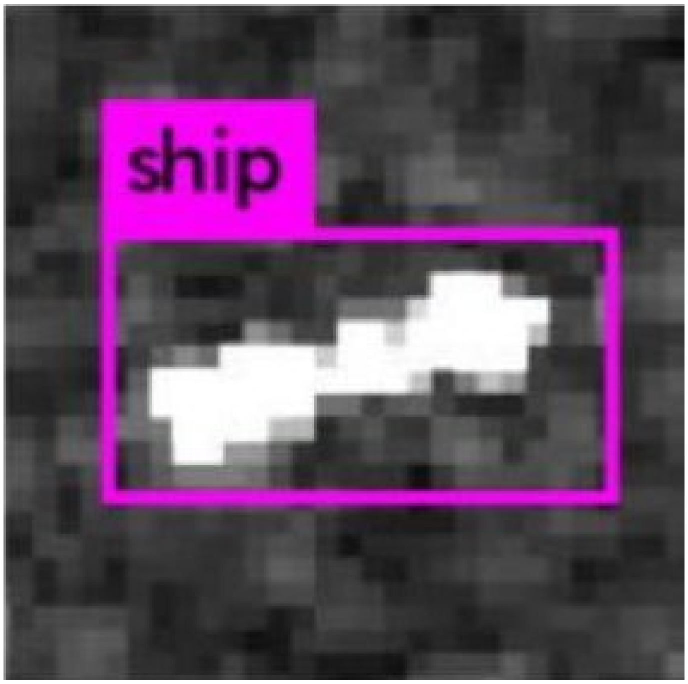


Figure 1: Sample annotated image

The following are two examples of images retrieved from the dataset. The images had already undergone a form of regularization, so it was not necessary to implement any further alteration in this regard.

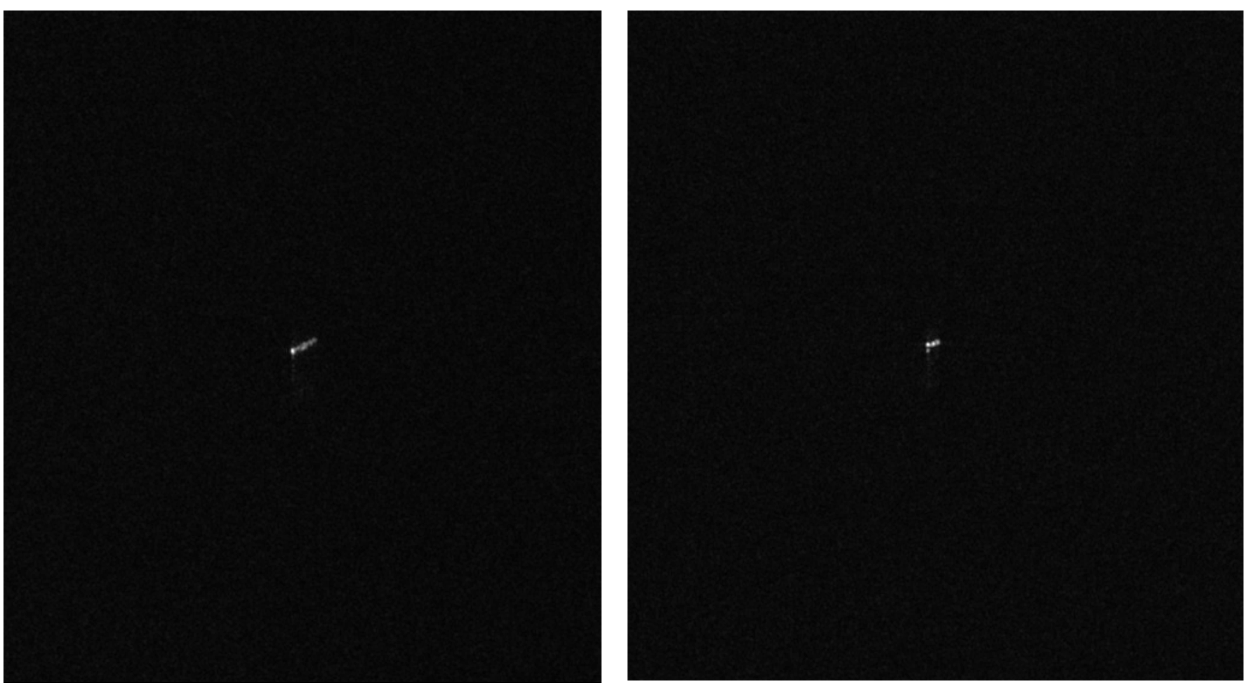


Figure 2: Two images containing ships

Algorithms, architectures/systems

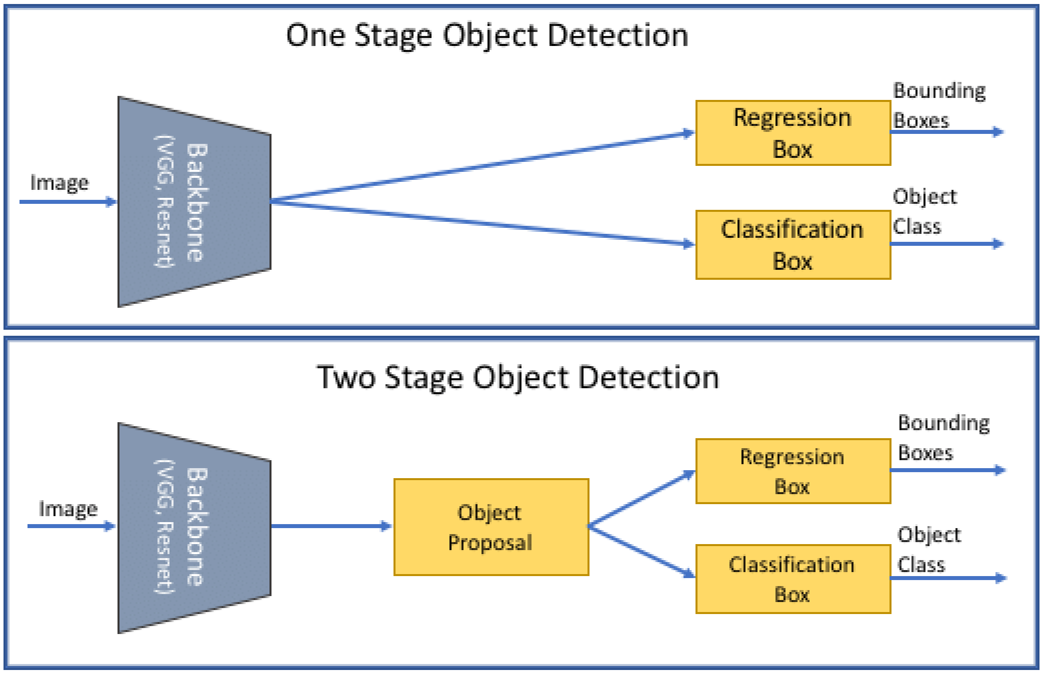


Figure 3: One-stage vs two-stage algorithms

The state-of-the-art object detection algorithms can be placed into two categories. The first one includes the algorithms that identify objects in two stages:

1) Find regions of interest
2) Classify objects and estimate their sizes with bounding boxes

In the second category of algorithms, object detection is achieved in one stage, leading to enhanced processing speed, but with the sacrifice of some degree of accuracy. In addition, one-stage methods are simpler in terms of architecture whilst two-stage include more complex pipelines. An example of a one-stage algorithm is SSD, which we employed for this project, and examples of two-stage algorithms are Region-based Convolutional Neural Networks (R-CNN), Fast R-CNN and Faster R-CNN. The following chart illustrates the relationship between some of the aforementioned algorithms in terms of speed and accuracy. In the light of those two factors, Fast R-CNN is generally inferior to Faster R-CNN and SSD. Faster R-CNN is generally superior in terms of accuracy but is slower compared to SSD. As can be concluded, Single Shot Detector achieves a good balance between speed and accuracy.

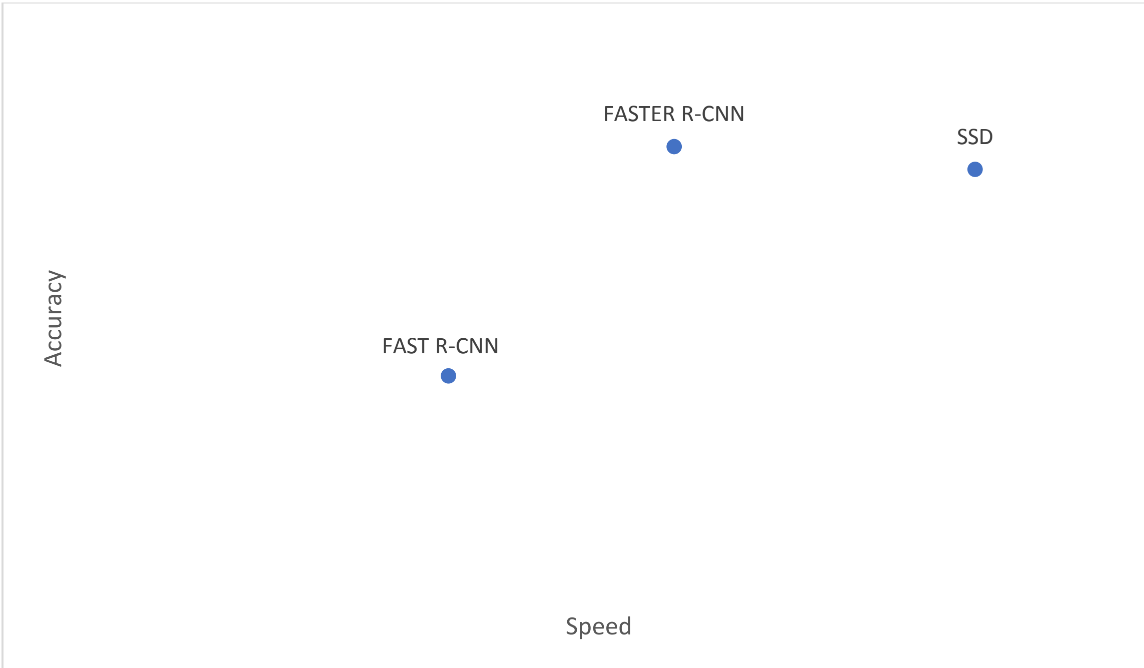


Figure 4: Comparison of various models in terms of speed and accuracy

SSD, also known as Single Shot MultiBox Detector, is inspired by the successfully CNN classifier architecture known as VGG-16, but rather than the fully connected layers in the end for object classification, the architecture has been revised to detect class-agnostic boundary boxes based on feature maps from different layers in the network. This approach is known as a single shot 13 approach. The illustration below shows how boundary boxes are detected from multiple layers in the network.

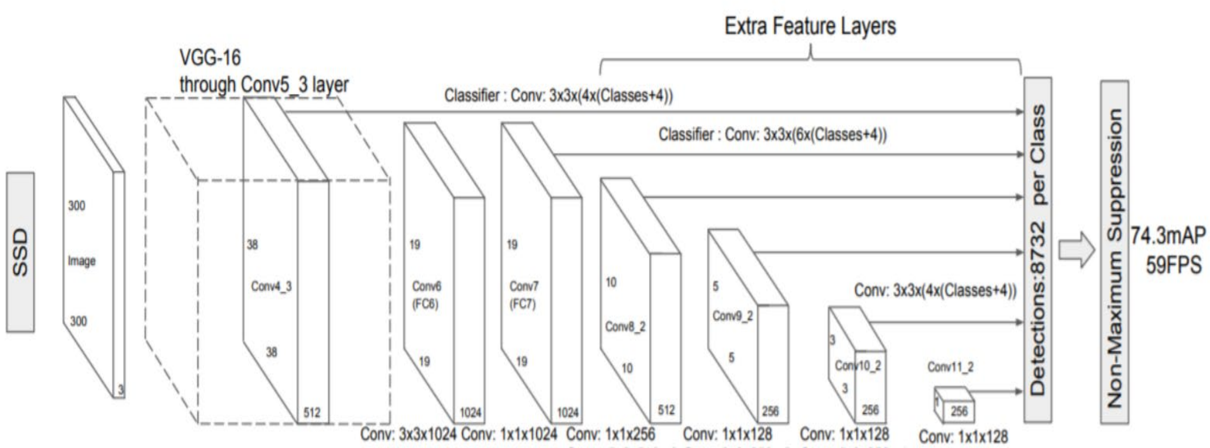


Figure 5: Graphical representation of how SSD works

Single Shot Detection needs an input image and ground truth boxes for each object during training. In a convolutional fashion, it evaluates a small set (e.g. 4) of default boxes of different aspect ratios at each location in several feature maps with different scales (e.g. 8×8 and 4×4). For the default boxes, it predicts both bounding boxes and classification probability. The model loss is a weighted sum between localization loss and confidence loss. The confidence loss measures how confident the network is at predicting the specific class. Categorical cross-entropy is used to compute this loss. Location Loss measures how far away the predicted bounding boxes are from the ground truth from the training set. L2-Norm is used to compute the location loss.

Experiments – Setup, Configuration

For the purpose of the first experiment, 1,193 observations were used for training, 299 for validation, and 99 for testing. Furthermore, various parameters were defined in accordance with our resources and project goals. Some of these parameters can be seen below:

- Number of training epochs: 200
- Batch size: 8
- Learning rate values: 0.00005;0.0001;0.00001
- Momentum for the optimizer: 0.9
- L2 normalization factor: 0.0005

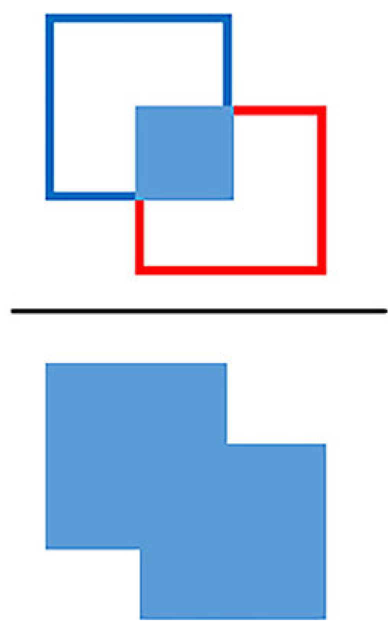
Results & Quantitative Analysis (incl. visualizations)

In order to evaluate our object detection model, we made use of various popular metrics such as mAP and IOU. We are going to explain them in more detail in the next paragraphs. Nonetheless, regardless of their differences, they all aim to answer a simple question:

How well the model performs on the ship detection task?

The first metric that we are going to analyze and which is of significant importance is IOU or Jaccard index as it is also known. **Intersection Over Union (IOU)** determines the difference between ground truth annotations and predicted bounding boxes. The following figure illuminates how it is calculated mathematically and what this calculation denotes visually.

$$IoU = \frac{area\ of\ intersection}{are\ of\ union} =$$



In object detection tasks, the metric is usually used as a threshold with its value inputted as a requirement at the beginning of the tasks. While the model is executed, it predicts multiple bounding boxes and removes those that do not satisfy the threshold. Two popular values for this metric that we are also using in this project are 0.50 and 0.75.

Precision is defined as the number of true positives divided by the sum of true positives and false positives:

$$precision = \frac{TP}{TP + FP}$$

Recall is defined as the number of true positives divided by the sum of true positives and false negatives. The sum is not anything other than the number of ground-truths.

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

The weighted harmonic mean of precision and recall or **F1-Score** is calculated as:

$$F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{(\text{precision} + \text{recall})}$$

This metric combines information from precision and recall so as to provide a more comprehensive result.

The final metric is called mean Average Precision or **mAP**. In our case, we can call it simple AP because with have only one class. Average Precision can be also interpreted as the area under the Precision-Recall curve. Because it incorporates the trade-off between precision and recall, it is usually preferred over the other metrics.

The following two graphs illustrate the magnitudes of the four metrics for two different thresholds in the testing dataset. Taking into account the first graph, we can conclude that the metrics take high values close to 90% indicating that the model exhibits robust ship identification capability. In particular, mAP takes the value 90.1%. In the case of IOU 0.75, as depicted in the second graph, the metrics report lower values. This is expected because as the threshold increases, only instances with large intersections are retained.

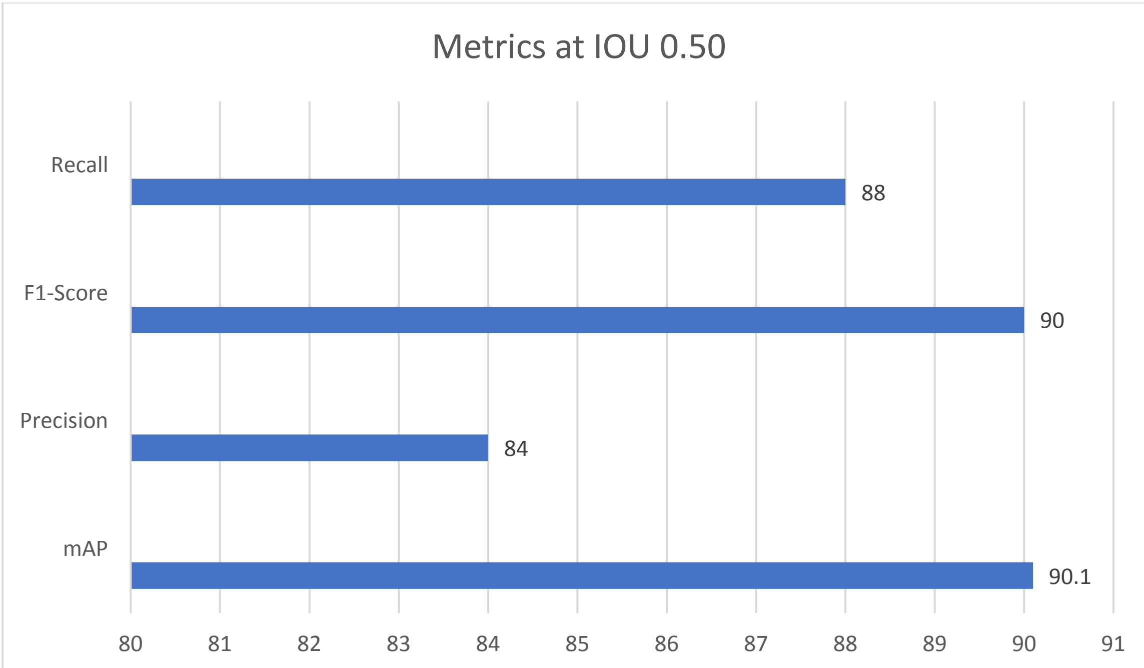


Figure 6: Metrics at IOU 0.50

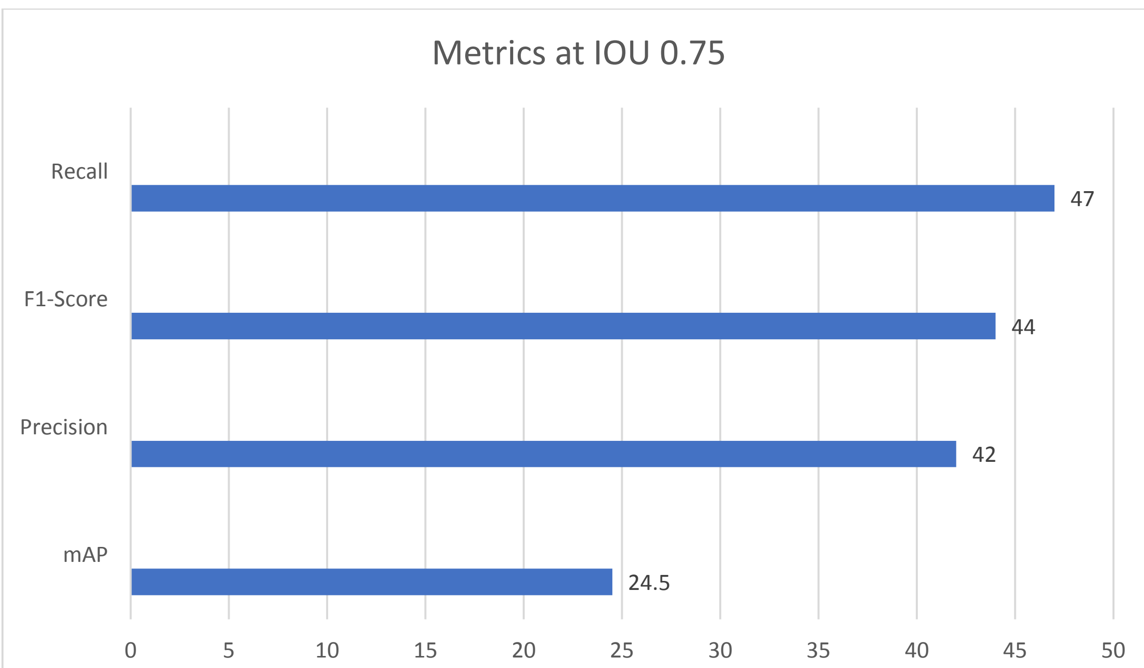


Figure 7: Metrics at IOU 0.75

Discussion, Comments/Notes and Future Work

Even though our model exhibited strong results, it does not mean that there is no room for improvement. The dataset that we used covered a specific geographic area and was limited in size. Expanding the size of the dataset and increasing the number of geographic areas can lead to a stronger model that can generalize well in many different situations. In an effort to ameliorate the model, we can compare it with other object detection algorithms, which incorporate new forms of architecture. This comparison can reveal weaknesses and areas of possible improvement. Another future work may relate to the classification capability of the model. In other words, we can turn the problem from binary classification to multiclass classification, identifying not only the existence of a ship but also its type.

Members/Roles

The project team consists of Adam Bellos, Alexandros Ziou, and Antonis Nakos, three students with bachelor degrees in mathematics, finance, and economics respectively. Adam comes from a more technical background and has previous experience with machine learning, while the other two are more business-oriented. Taking into account the strengths of each individual team member, tasks were divided with the aim of producing the best possible outcome. For that reason, Adam undertook the bulk of the technical part and guided the rest through the process so that they could also be able to contribute in certain technical aspects. In the meantime, Antonis and Alexandros also concerned themselves with the business aspect of the project, researching the Maritime sector, identifying the current landscape and the need for Ship Detection Systems, organizations that would be interested in such a model, etc. as well as theoretical aspects regarding the project, including the study of previous work on object detection to identify the optimum kind of model the team should opt for in order to proceed with the technical part. Finally, it should be stated that Adam took the responsibility of compiling the vast majority of the code, while Alexandros and Antonis were assigned the task of producing this report, having a well-rounded grasp of the overall project.

Bibliography

1. Hui, J. (2020, December 15). Ssd object detection: Single shot multibox detector for real-time processing. Medium. https://jonathan-hui.medium.com/ssd-object-detection-single-shot-multibox-detector-for-real-time-processing-9bd8deac0e06.
2. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.