Comparative Study of One-Stage Detection Techniques with SAR images

Philip Bellos*, Theodore Betsas**,Emmanuel Bratsolis***, Eleni Charou****
*National and Kapodistrian University of Athens, Athens, Greece, fbellos@phys.uoa.gr
**National Technical University of Athens, Athens, Greece, betsasth@gmail.com
***University of West Attica, Athens, Greece, ebrats@uniwa.gr
****NCSR Demokritos, Athens, Greece, exarou@iit.demokritos.gr

Abstract: Synthetic Aperture Radar's (SAR) monitoring capabilities regardless of clouds or daylight, render it a valuable source of data and a superior option to optical satellite images for the demanding task of ship detection[5]. Traditional deep learning techniques are computationally intensive and inefficient, while two-stage detectors' need for handling different components due to sparse detection are proven not to be the ideal solution. In order to accelerate SAR imagery analysis and achieve realtime detection, three one-stage ship detectors namely Single Shot Multibox Detector (SSD), YOLOv3 and YOLOv4 are studied and compared in this paper. The detectors are trained by optimizing classification-loss and localization-loss simultaneously. To this end, two datasets for training and evaluating the ship detectors are used: an existing public SAR ship detection dataset and the Demokritos SAR Ship Dataset (DSSD), which we created via automatic methodology that we implemented. For the DSSD we additionally applied a pseudolabelling technique exploiting the detection results of the existing dataset. Experimental results illustrate that the most recent YOLOv4 model achieves a superior detection performance in both datasets when compared to the SSD and YOLOv3 models, while the reasons behind these results are identified and examined.

Existing Dataset[1]:

Contains 1596 ship images created from two main subsets of data from Sentinel-1 and RADARSAT-2.
First subset: 43 Sentinel-1 Extra Wide (EW), swath acquisition mode and Ground Range Detected (GRD) images. 2 different resolutions: GRDH(EW) images with a resolution of 50 x 50 m and pixel spacing of 25 × 25m in range and azimuth respectively and GRDM(EW) images with a resolution of 93 × 97m and 40 x 40m pixel spacing.
Second subset contains 3 RADARSAT-2 ScanSAR Narrow (SCNA) imagery of the non SLC type (intensity only). The imagery has a resolution of 81×30 m and pixel spacing of 25×25 m (in range and azimuth respectively).
Therefore, the ship image dataset was created by a total of 46 SAR images.

Demokritos Sar Ship Dataset (DSSD)[2]:

For the DSSD, images from the Sentinel 1 mission were selected and from the available Sentinel 1 products, the GRD IW type. Finally, it contains 2285 ship patches with square pixels, a spatial analysis of 20*22 m, pixel spacing of 10*10 m and a reduced speckling. The size of each patch is 512*512 pixels.

Automated Process for DSSD construction:

The Demokritos SAR Ship Dataset is collected and created via an automatic procedure that we are introducing. The steps of the process's implementation are described below:

1. The main part (red section), receives the dates and the geographical coordinates of the point of interest. Then, all the images in the specified dates, that contain the point of interest, are downloaded.
2. Afterwards, the ESA's ocean object detection algorithm runs via the command line (green section) and produces the ShipDetections.csv (SD.CSV) file, that contains the detected ship's coordinates.
3. This file is processed, (blue section) and the patches are created by cropping the downloaded images in an area of 512*512 pixels around the detected coordinates. For the dataset creation points in the middle of the Mediterranean sea were selected as points of interest, to avoid miss-detections due to land-sea masking.

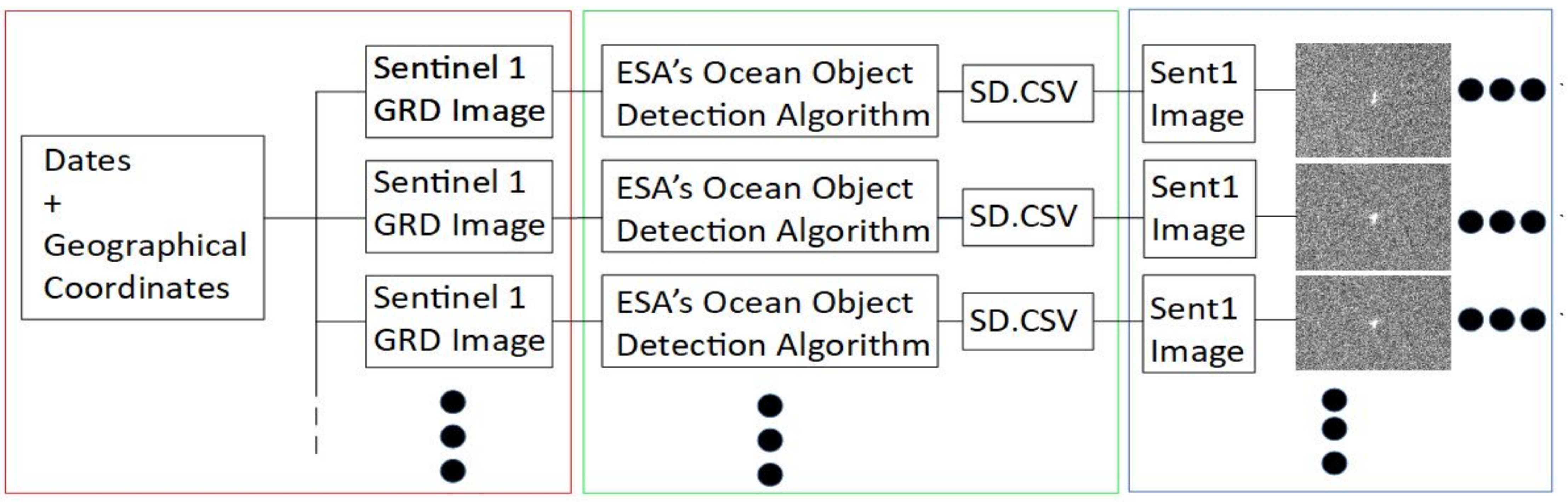

Figure 1. DSSD automatic collection and construction method

Preprocessing:

The annotation of all the 1596 SAR images of the existing dataset was done manually.with the open source YOLO Mark tool[12]. For the DSSD a different, more efficient and less time consuming method was applied.
We first choose the best performing model(YOLOv4) to be used as an inference model to generate pseudolabels For the DSSD. We end up with annotations for all of the SAR images of the dataset in YOLO format. At this stage however, the dataset contains ship images with faulty annotations due to False Positive and False Negative detections, but also from True Positive ones that have a relatively low IoU score. Those falsely annotated images are semi-automatically discarded, since the nature of our dataset enables such a task (DSSD consists of images that contain 1 to 3 ships). The elimination of those annotations is based on the memory consumed by each file containing the latter. If the file consumes no memory then no detection has been made, leading to discard it. If the memory corresponds to more than the ships that were supposed to be in the image, then that image is forward passed to YOLOv4 again but now with an IOU set at 0.65, keeping the most confident detections. Finally, with the dataset almost clean, a manual supervision is done on the images where the discrepancy between memory and number of ships remains even after the thresholding.

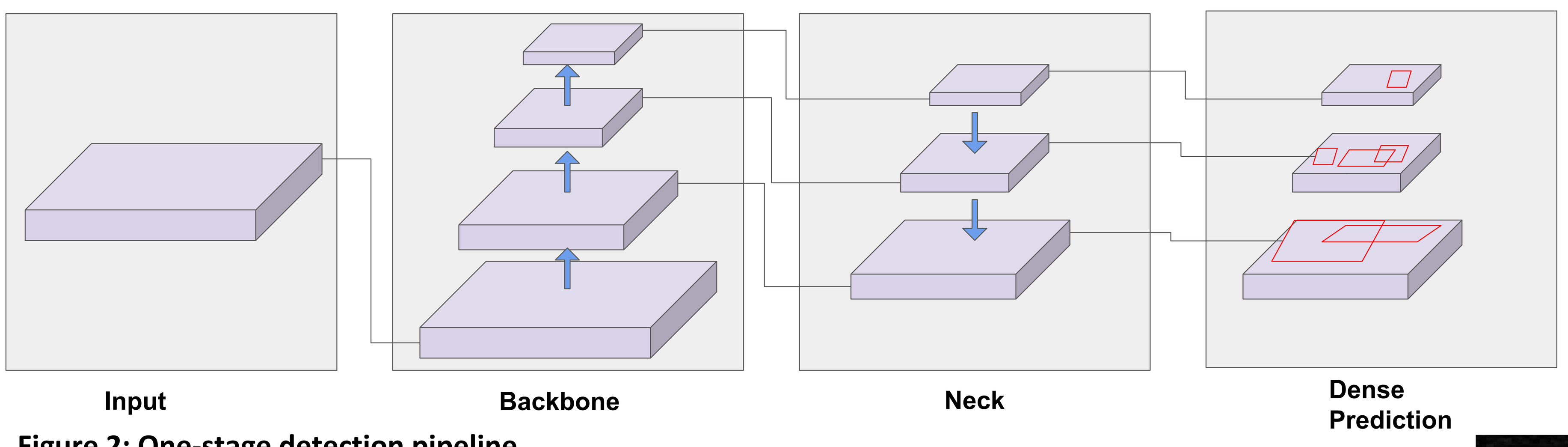

Figure 2: One-stage detection pipeline

In this work, one stage detection models,namely SSD[4],YOLOv3[3] and YOLOv4[6] are selected to perform ship detection on SAR imagery. While 2-stage detectors are composed of several correlated stages, including region proposal generation, feature extraction with CNN,classification and bounding box regression, which are usually trained separately, 1-stage ones are able to predict the object class directly from the densely pre-designed candidate boxes due to the head of those models that applies dense prediction. Consequently, they battle this bottleneck of handling different components and enhance performance. Detection time and performance are crucial for our task, as the applications of ship detection need to be real time and extremely accurate.
We compare SSD,YOLOv3 and YOLOv4 on both datasets. Each result is the mean of two separate trainings with different train-validation-test split sets for assessing the models' performance more accurately. The information involved includes the test results for IoU set at 0.50 and 0.75,and the backbone network .
As can be seen from the tables YOLOv4 performs better than SSD and YOLOv3 when Average Precision is considered while for the other metrics YOLOv4 and YOLOv3 show a similar performance. With YOLOv4 demonstrating overall better results we continue our experiments further with it on larger images, specifically on 576 × 576 images.

Model	Backbone	Dataset	IOU@0.50				IOU@0.75			
			mAP	Precision	F1-Score	Recall	mAP	Precision	F1-Score	Recall
SSD	Modified VGG-16	Existing	90.1	84	90	88	24.5	42	44	47
		DSSD	88.9	83	89	89	26.7	45	46	47
YOLOv3	Darknet53	Existing	91.98	84	87	92	33.76	48	50	52
		DSSD	92.84	86	89	89	30.22	45	47	49
YOLOv4	CSPDarknet 53	Existing	92.94	87	89	92	35.11	52	53	54
		DSSD	92.89	86	89	90	32.08	50	50	51

Table I: Results on Test set

Model	Dataset	IOU@0.50				IOU@0.75			
		mAP	Precision	F1-Score	Recall	mAP	Precision	F1-Score	Recall
YOLOv4	Existing	91.67	83	87	91	35.57	49	51	54
	DSSD	96.53	92	93	95	31.49	47	48	48

Table II: Results on Test set(576x576 images)

Model	Time(ms)	BFLOPS
SSD	52.63	-
YOLOv3	34.72	98.923
YOLOv4(512)	24.30	90.226
YOLOv4(576)	28.26	114.92

Table III: Detection time and BFLOPS

We can see that YOLOv3 and YOLOv4 achieve near real-time detection times with YOLOv4 being the faster one. Our YOLOv4 ship detector performs even better on both datasets when trained on those larger images. However, when trained in larger images while it can be seen to perform better, that improvement comes at the expense of detection time and Billion Floating Point Operations (BFLOPS).

Why YOLOv4 works better?

The Mosaic data augmentation. It combines four images in one during training allowing batch normalization to calculate activation statistics from 4 different images on each layer. We train our model with and without mosaic enabled, observing an increase in mAP of 0.3%.

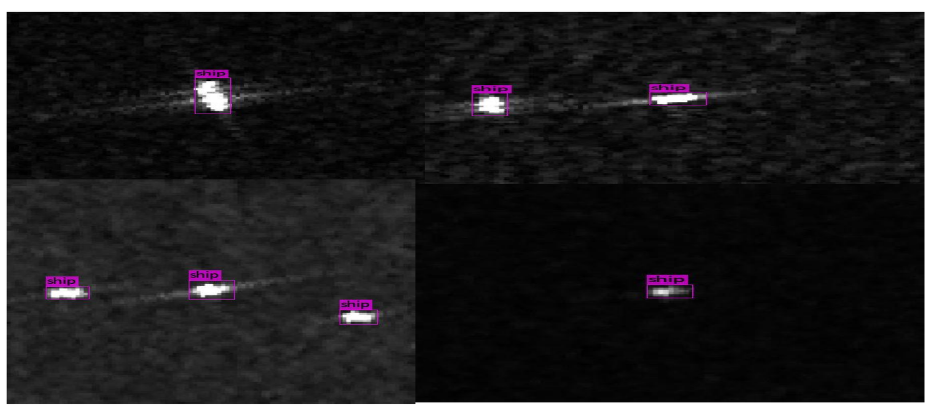

Figure 3: Mosaic Data Augmentation

Dropping continuous regions of the input image (left) is more effective when compared to just dropping out random activations(right).Removing important semantic information (e.g. altering the shape of the ship, remove the whole ship)forces the network to learn features for the classification of the input ship image.

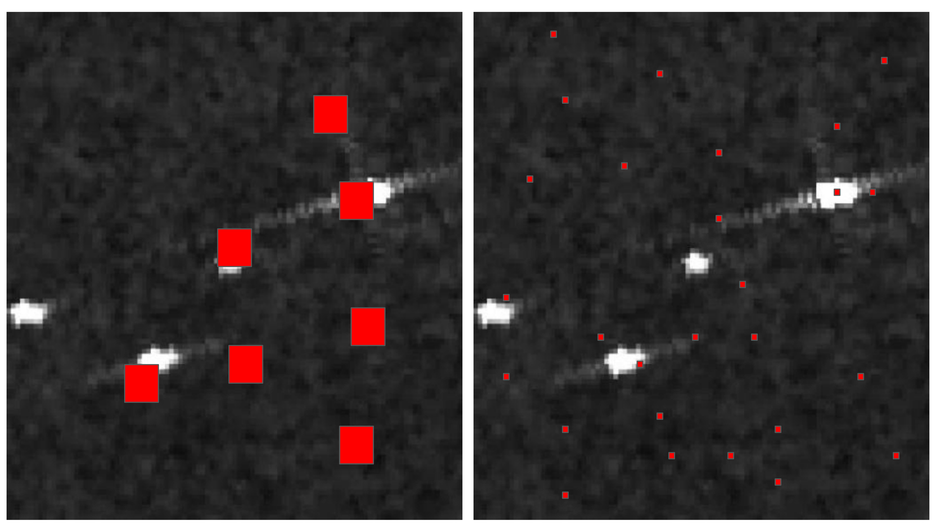

Figure 4: Dropblock (left), Dropout(right)

The Mish activation function. Mish works better than ReLU in many deep networks acro challenging datasets. Mish is a novel smooth and non-monotonic neural activation functi which can be mathematically defined as: $f(x)=x\tanh(\text{softplus}(x))$.

The CIoU loss. It takes into consideration the overlapping area, the distance between center points and aspect ratio. The figure compares bounding boxes' localization and detection for the same image with different types of IoU losses. We can notice that YOLOv4's loss outperforms YOLOv3's.

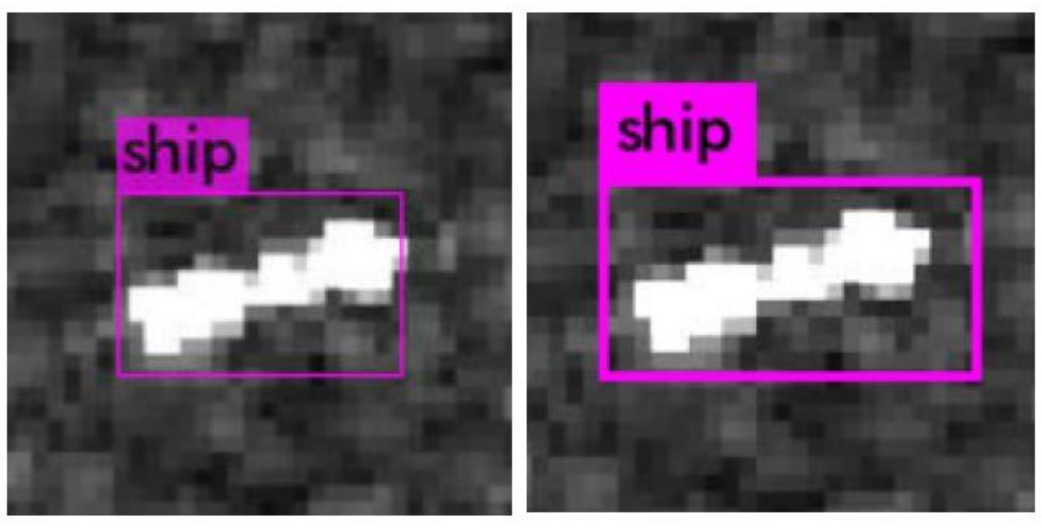

Figure 5: CIoU loss of YOLOv4(left), IoU loss of YOLOv3(right)

Future work:

Numerous applications are related to the automatic detection of ships. In cooperation with ship tracking applications, information can be obtained, about ships located near illegal discharges, such as oil spills, about ships above or near underwater antiquities or "black" ships. Furthermore, with various techniques[7], ship gas emissions can be recorded. Those recordings in combination with our ship detection, can provide additional valuable information.

References:
[1]]C. P. Schwegmann and W. Kleynhans and B. P. Salmon and L. W. Mdakane and R. G. V. Meyer "A SAR ship dataset for detection, discrimination and analysis" IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016 jul, (104--107)
[2]Theodore Betsas,Philip Bellos, Emmanuel Bratsolis,Eleni Charou,Dataset creation and ship detection using SAR images with deep learning techniques,2nd Workshop on Remote Sensing and Space Applications,Geological Society of Greece,February 26,Athens,Greece,2020.
[3]Joseph Redmon, Ali Farhadi. YOLOv3: An Incremental Improvement,arXiv:1804.02767
[4] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, SSD:Single Shot MultiBox Detector, arXiv:1512.02325
[5] E. Charou, E. Bratsolis, S. Gyftakis, T. Giannakopoulos and S.Perantonis, "Use of Sentinel-1 data for Maritime Domain Awareness: Preliminary results,"2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA), Corfu, 2015, pp.1-4, doi: 10.1109/IISA.2015.7388021
[6] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection,arXiv:2004.10934
[7] Iosifidi-Maria (2020) 'Study of exhaust gas emissions from conventional merchant vessels' Diploma Thesis, School of Naval Architecture and Marine Engineering, National Technical University of Athens, Athens