Multiple Vessel Detection and Tracking in Harsh Maritime Environments

Diogo Ferreira Duarte
INESC TEC
Faculty of Engineering
University of Porto
up201405740@edu.fe.up.pt

Maria Inês Pereira
INESC TEC
Faculty of Engineering
University of Porto
maria.i.pereira@inesctec.pt

Andry Maykol Pinto
INESC TEC
Faculty of Engineering
University of Porto
andry.m.pinto@inesctec.pt

***Abstract*—Recently, research concerning the navigation of Autonomous Surface Vehicles (ASVs) has been increasing. However, a big scale implementation of these vessels is still held back by a plethora of challenges such as multi-object tracking. This article presents the development of a tracking model through transfer learning techniques, based on referenced object trackers for urban scenarios. The work consisted in training a neural network through deep learning techniques, including data association and comparison of three different optimisers, Adadelta, Adam and SGD, determining the best hyper-parameters to maximise the training efficiency. The developed model achieved decent performance at tracking large vessels in the ocean, being successful even in harsh lighting conditions and lack of image focus.**

***Index Terms*—ASV, Multiple Object Tracking, Object Detection, Machine Learning, Data Augmentation, Deep Learning**

I. Introduction

The inclusion of Artificial Inteligence (AI) in the maritime environment is fairly recent. However, the progress done with autonomous navigation of terrestrial vehicles, among other areas, shows a lot of promise for maritime autonomous navigation in the near future [1]. In the context of Autonomous Surface Vehicles (ASVs), tracking systems are commonly used to avoid collisions. It is important to identify each detected object in the vehicle's surroundings and predict their movement. Different vessels may also present different behaviours, so their individual characteristics (such as shape and size) should also be studied. Furthermore, the sea doesn't constitute a flat and stable surface to place a stationary observer. Having a dynamic observer inserts unpredictable tilts that introduce a bigger challenge in trajectory prediction which is relevant for multiple object tracking. Multiple object tracking systems on water is a subject studied by few, without solid results. The work researched showed great promise in obstacle detection and tracking but was reliant on information from on-shore systems or previously collected data from the same testing environment. In 2017, Tan *et al.* [2] developed a system capable of detecting obstacles using a digital camera suitable for river navigation. However, location was determined by longitude coordinates plotted using *Google Earth*. In 2019, Freire *et al.* [3] published their model for obstacle avoidance combining a radar for shore detection, an AIS to relay location information of the bigger vessels that have an AIS transceiver, while using LiDAR to detect smaller vessels and debris. GPS

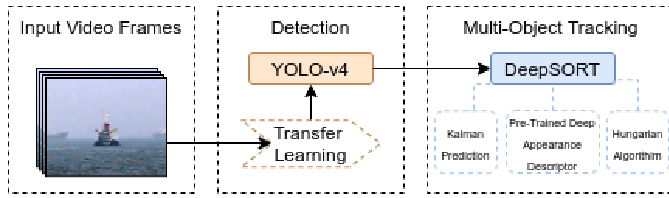

Fig. 1. Model Diagram of the architecture. Image frames are fed to the detector module, which classifies every object in each frame and processes their bounding boxes, feeding that information to the tracking module.

and IMU data are relied on for extraction of the own ship's position and orientation. Unlike these approaches, the work described in this article aims to develop a robust camera-based detection and tracking system that can be used with on-board hardware. Its development resorts to transfer learning techniques, fine-tuning of training parameters and optimiser selection, on available object detectors and trackers that perform best under maritime environments, in order to achieve successful vessel tracking at sea.

II. Multi-Object Tracking Architecture

This section aims to exhibit the research done selecting the backbone of the adapted detection and tracking model, while presenting a brief diagram of its architecture. As represented in figure 1, we opted for a modular approach which allows for better control of the system, improving the detections first to then focus on the tracking module.

A. Detector Selection

The scarcity of annotated maritime datasets for object detection and classification deemed unfeasible a similar performance comparison among popular object detectors with open-source implementations. However, such a comparison exists with the MSCOCO [4] dataset, which ranks YOLOv4 [5] as the current state-of-the-art in real-time object detection. This detector is an improvement of the original YOLO [6], adding a backbone network, feature aggregation and activation functions. Developed by Chien-Yao Wang *et al.* in 2020, it uses the CSPDARKNET53 [5] neural network as backbone, combined with traits such as feature aggregation, a "Bag-of-Freebies" such as data augmentation, and a "Bag-of-Specials"

978-0-692-93559-0 ©2021 MTS

that makes use of activation and loss functions, making it now more efficient and suitable for single GPU training.

B. Tracking Model Comparison

With ASVs, it is important to consider tracking solutions with online methods in order to track in real life scenarios. A set of popular object trackers with open-source available implementations were evaluated using as input videos from the On-Board section of the Singapore Maritime Dataset [7]. From the data available, those videos represented the most difficult conditions commonly found under maritime environments such as rippling, fog, camera focus and exposure issues. DeepSORT [8] achieved the best performance, managing to track objects in real time, including fast boats, while failing to associate to the same object only in some situations of unnatural camera tilt or lack of focus. This state-of-the-art tracker aims to correct the association metric issue from the original SORT [9] with a more informed feature vector that combines motion and appearance information. The feature vector is obtained applying a CNN to compute bounding box appearance descriptors for the type of objects intended to track. Its approach still uses a combination of the Kalman Filter and the Hungarian algorithm to predict and associate data respectively.

III. Transfer Learning

The following subsections describe the transfer learning process applied to the object detector, including the conversion of the previously mentioned datasets to its network format and the fine-tuning of its training parameters in order to achieve the best overall average precision in maritime scenarios.

A. Maritime Datasets

Data concerning maritime objects with ground-truth and proper classification is not widely available. For the training module, the Singapore Maritime Dataset [7] provides on-shore and on-board recorded videos with matrix files of the bounding boxes of every object in every frame of every videos. For training of the classifier in the tracking module, the MARVEL Dataset [10] provides a set of over 100k images of different vessels distributed over 38 classes.

B. Data Augmentation

The darknet input format in YOLO-v4 requires a dataset composed of images. In order to use the Singapore dataset , frames were extracted each n seconds from On-Shore videos and a $.txt$ file with the classification and bounding box of each object in the frame was generated, totalling 1828 images. The expansion of this dataset was achieved with data augmentation techniques as lowering the n value would generate a dataset with too many similar images. Common augmentation techniques such as image rotations are already included in the "Bag-of-Freebies" part of the YOLO-v4 training process. Instead, our augmentation focuses on applying image filters such as histogram correction and sharpening with the purpose of minimizing the exposure issues created by the maritime environment. Figure 2 illustrates the image filtering.

Fig. 2. Visual comparison of one of the applied filters to the original frame. The objects in the image are better exposed and contain more defined edges, allowing the detector to better identify their shape and colours.

C. Optimisers

Finding the best numerical solution to a given problem is an important part of many branches in mathematics, and machine learning is no exception. Optimisers tie together the loss function and model parameters by updating the model in response to the output of the loss function. For this work, we compare the performance of three popular optimisers: the original Stochastic Gradient Descent (SGD) [11], Adaptive Moment Estimation (Adam) [12] which is a well referenced improvement of the original SGD, and Adadelta [13], a more robust extension of the popular Adagrad [14].

D. Parameter fine-tuning

The generated images were then divided into a training and validation dataset with a 9:1 ratio. The on-board videos are used as real footage to test the whole tracking system and provide results as they contain harsher observer conditions. A training exercise with the same conditions (batch size, image size, number of epochs) was made for each optimiser, varying only the initial learning rate, $lr0$, and momentum parameters, each for 3 arbitrary values. The training results are then plotted and compared to determine which optimiser works best for the provided dataset and which parameters potentiate the performance of the best optimiser.

IV. Results

The metric used to compare the accuracy of the different training exercises was the Mean Average Precision (mAP). For the first batch, only the mAP with Intersection over Union (IoU) of more than 50% was considered. This metric calculates the mean of the precision metric (true positive detections over the total positive detections value) for each class of objects in the dataset. For the second and third batches we also included the mAP[.5:.95] which iterates the average of all mAP with IoU bigger than 50% in intervals of 5%. Measurements of the mAP are displayed in tables I, II and III. Overall, the Adam and Adadelta optimisers achieved consistent over 80% mAP@.5 with the learning rates of 0.001 and 0.1 respectively. With only one result ($lr0 = 0.05$ and $momentum = 0.8$) achieving similar performance, the SGD optimiser was deemed less fit for this specific training scenario.

Of the 27 training exercises, the top 2 of each optimiser are plotted in figure 3. The best result was achieved with the Adam optimiser with the parameters $lr0 = 0.001$ and $momentum = 0.8$.

TABLE I
MEASURED MAP@.5 FINAL RESULTS USING THE ADAM OPTIMISER. BEST RESULT IN BOLD.

Adam		Momentum		
		0.8	0.95	0.5
lr0	0.05	1.010e-5	0.2769	4.510e-5
	0.001	**0.8498**	0.8295	0.8165
	0.1	0.1166	0.1033	4.170e-5

TABLE II
MEASURED MAP@.5 FINAL RESULTS USING THE ADADELTA OPTIMISER. BEST RESULT IN BOLD.

Adadelta		Momentum		
		0.8	0.95	0.5
lr0	0.05	0.7249	0.7538	0.6611
	0.001	0.0945	0.1388	0.0782
	0.1	**0.8309**	0.824	0.8114

Through analysis of figure 3, the top three results come very close. Furthermore, the training wasn't extensive enough to stale out the evolution curves. While the SGD optimiser is clearly inferior, it is still hard to tell whether Adam or Adadelta would come out on top. To rule out the uncertainties, a second training exercise was done, testing out the best $lr0$ and $momentum$ combinations for each of the top two results.

The figure 6 shows the results of 40 epochs of training. Adadelta ended up coming on top after the 10th epoch, remaining superior through further stages. In this exercise we were able to achieve a 17.5% increase in mAP@.5 compared to the previous best Adadelta result, while managing a 14.9% boost with the Adam optimiser. Both values, however, are awfully close with a less than 1% difference, meaning that there isn't a significant difference between detections with both optimisers.

In order to better compare Adam and Adadelta, we observed the evolution of the mAP[.5:.95] metric throughout the training process. This metric better exhibits the quality of the detections, meaning that higher values will point to detections with a higher IoU. The figure 7 displays the evolution of the mAP[.5:.95] metric over epochs, illustrating a 3.8% improvement between the two optimisers.

Furthermore, in order to explore the full potential of the winning configuration, a third train was done until the mAP@.5 evolution stabilized. Through this last training, the mAP@.5 values did not increase further than 1%. However, the figure 7 illustrates a mAP[.5:.95] evolution of 16.9% over the last 40

TABLE III
MEASURED MAP@.5 FINAL RESULTS USING THE SGD OPTIMISER. BEST RESULT IN BOLD.

SGD		Momentum		
		0.8	0.95	0.5
lr0	0.05	**0.8127**	0.3363	0.7377
	0.001	0.4025	0.6973	0.2738
	0.1	0.3604	0.4150	0.4995

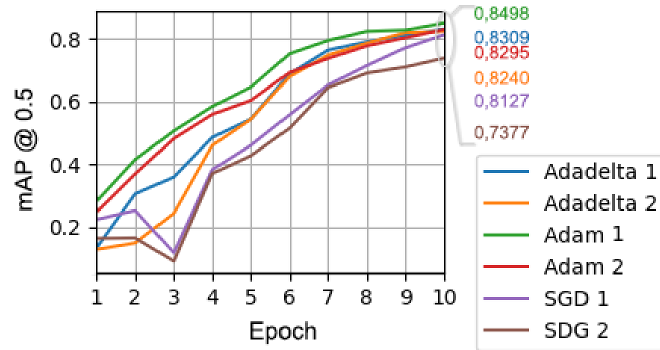

Fig. 3. Mean Average Precision of detections with IoU over 0.5.

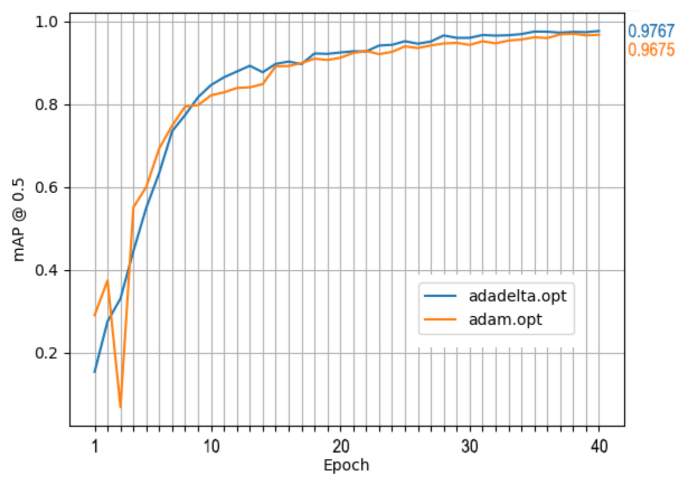

Fig. 4. Mean Average Precision with IoU over 0.5 comparison between the optimal Adadelta and Adam optimised configurations.

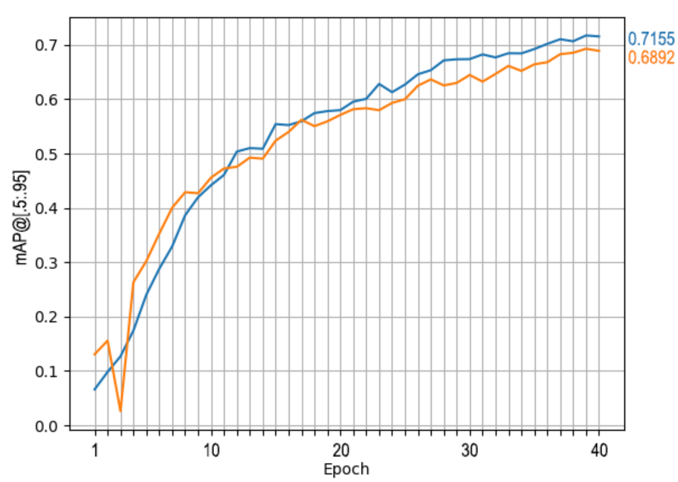

Fig. 5. mAP[.5:.95] comparison between the optimal Adadelta and Adam optimised configurations.

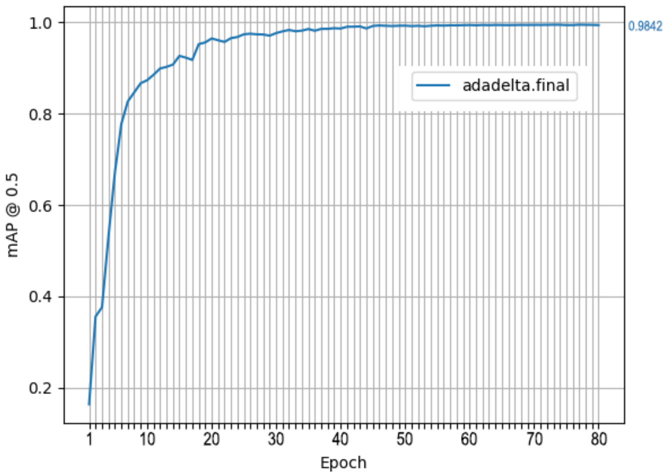

Fig. 6. Mean Average Precision with IoU over 0.5 results of a densely trained Adadelta setting.

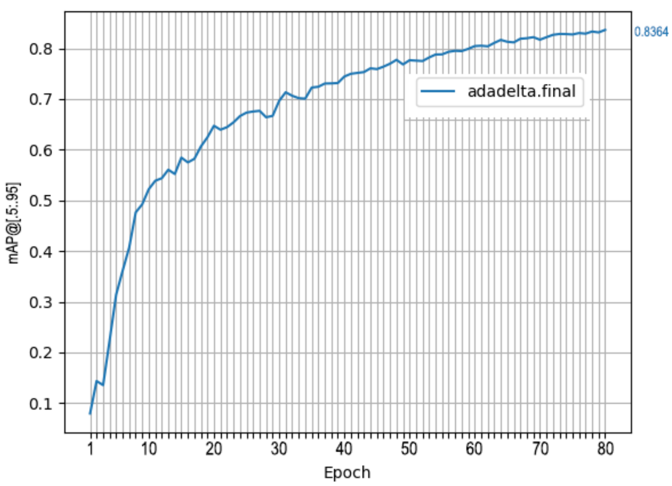

Fig. 7. mAP[.5:.95] evolution in a densely trained Adadelta setting.

epochs.

A. Tracking Results

Figure 8 shows successful tracking of a vessel using a re-trained YOLO-v4 detector model in conjunction with an also re-trained DeepSORT tracker model. The bounding box isn't accurately around the object, evidencing a need of further optimisation in training the neural network. An interesting option would be an optimisation of the training length, measuring the longest possible drill that increases the mAP[.5:.95] metric while keeping the over-fitting values low. With the available testing machine powered by an Nvidia GTX 960M, the model was able to successfully track up to 5 vessels at the same time while running on an average of 18 frames per second.

V. Conclusions & Future Work

In this article, Deep Learning models for detection and tracking were studied and compared. The selected modules

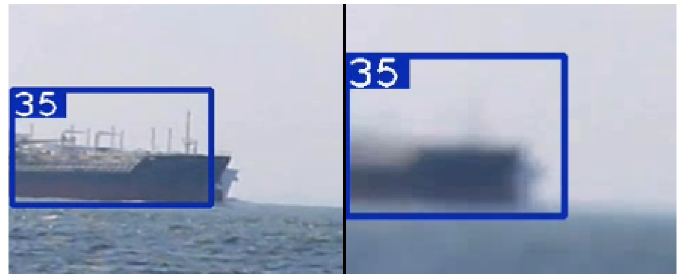

Fig. 8. Successful tracking of a vessel under harsh observer conditions. The number in the top left corner indicates the invidivual object's label.

were adapted through transfer learning to maritime environments and put to the test. The resulting system is able to overcome harsh observer conditions and successfully detect vessels at open sea. To achieve better tracking results, future work on reinforcing the tracker's prediction is needed in order to deal with excessive camera tilts provoked by unpredictable ocean ripples. Furthermore, the datasets available for this particular scenario are scarce and without enough diversity. The construction of a better suited dataset for detection and tracking in maritime environments can also be the basis of better performance for future work in the subject.

References

[1] F. Ben Mas, D. Piccolo, L. Cobianchi, L. Edvinsson, G. Presch, M. Massaro, M. Skrap, A. di Tor Vajana, S. D'Auria, and C. Bagnoli, "The effects of artificial intelligence, robotics, and industry 4.0 technologies. insights from the healthcare sector," (Reading, UK), pp. 88 – 95, 2019//.

[2] C. S. Tan, M. R. Arshad, and R. Mohd-Mokhtar, "River navigation system using Autonomous Surface Vessel," *USYS 2016, IEEE 6th International Conference on Underwater System Technology: Theory and Applications*, pp. 13–18, 2017.

[3] D. Freire, J. Silva, A. Dias, J. M. Almeida, and A. Martins, "Radar-based target tracking for obstacle avoidance for an autonomous surface vehicle (asv)," in *OCEANS 2019 - Marseille*, pp. 1–6, 2019.

[4] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Lecture Notes in Computer Science*, vol. 8693 LNCS, pp. 740–755, 2014.

[5] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020.

[6] 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, December 2016.

[7] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "Video Processing From Electro-Optical Sensors for Object Detection and Tracking in a Maritime Environment: A Survey," Tech. Rep. 8, 2017.

[8] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," *Proceedings - International Conference on Image Processing, ICIP*, vol. 2017-Septe, pp. 3645–3649, 2018.

[9] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proceedings - International Conference on Image Processing, ICIP*, vol. 2016-Augus, pp. 3464–3468, 2016.

[10] E. Gundogdu, B. Solmaz, V. Yucesoy, and A. Koc, "Marvel: A large-scale image dataset for maritime vessels," in *Asian Conference on Computer Vision (ACCV)*, 2016.

[11] J. Kiefer and J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function," *The Annals of Mathematical Statistics*, vol. 23, no. 3, pp. 462 – 466, 1952.

[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[13] M. D. Zeiler, "Adadelta: An adaptive learning rate method," 2012.
[14] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, p. 2121–2159, July 2011.