2022 International Conference on Unmanned Aircraft Systems (ICUAS)
Dubrovnik, Croatia. June 21-24, 2022

2022 International Conference on Unmanned Aircraft Systems (ICUAS) | 978-1-6654-0593-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICUAS54217.2022.9836137

Automatic detection of rescue targets in maritime search and rescue missions using UAVs

Luís Gonçalves[1] and Bruno Damas[2]

***Abstract*— Unmanned Aerial Vehicles (UAVs) can be an important resource when performing Search and Rescue (SAR) operations at sea, as this technology is fairly inexpensive when compared to traditional SAR approaches that use significant human resources and expensive air and naval assets, thus enabling the deployment of several UAVs simultaneously in these missions to perform rescue targets search in maritime environments.**

In order to maximize the usefulness of these UAVs in such operations, we propose a method which utilizes a state-of-the-art object detection network to perform real-time rescue target detection on-board the UAV, using standard RGB cameras, with minimal human intervention, thus enabling an increased vehicle autonomy and search range. Additionally, since the UAVs only relay the candidate images and locations that contain possible rescue targets, given by the onboard detector, it is possible to have several UAVs working in parallel that report back to a single human operator.

We have selected the YOLOv4-tiny detection network, pretrained in the COCO dataset, and retrained it to detect rescue targets at sea. For this purpose some datasets were recorded and annotated to simulate the presence of maritime rescue targets. The proposed approach has been validated on an independent test dataset, showing that it has good detection capabilities and thus providing convincing results regarding the use of UAVs with automatic target detection capabilities in SAR missions.

I. Introduction and Related Work

Travel through sea has long been a crucial part of the world economy, and nowadays around 90 % of traded goods are carried by sea: as a result, there is a significant number of ships at sea at any given time [1].

With such large number of ships around the globe it is not surprising the ocasional occurrence of shipwrecks and man overboard incidents (Fig. 1): this lead to the need for an ever evolving response capability in terms of Search and Rescue (SAR) operations at sea. The International Maritime Organization (IMO) and the International Civil Aviation Organization (ICAO) have jointly published the IAMSAR Manual [2], which provides guidelines for the organization of maritime SAR response.

This work was supported by the Portuguese national project VOAMAIS (02/SAICT/2017/31172)

[1]Luís Gonçalves is with CINAV, Centro de Investigação Naval, Portuguese Navy, Portugal vale.gonçalves@marinha.pt

[2]Bruno Damas is with CINAV and with ISR, Institute for Systems and Robotics, Lisbon, Portugal bruno.damas@escolanaval.pt

Fig. 1. Man overboard incident. Image taken from https://www.cdc.gov/niosh/topics/fishing/fallsoverboard.html.

In Portugal, most SAR operations count on aircraft and ship support in order to find and retrieve rescue targets, but this requires a lot of resources and manpower. The evolving technology in the field of unmanned aerial vehicles (UAVs) is an oportunity to improve the response capabilities of the Portuguese Navy and Air Force, with the deployment of less resources and manpower. With the use of UAVs we can greatly extend the covered search area and even automate some processes previously done by human operators. The use of UAVs can be done on three distinct levels:

1) In the simplest method, human operators perform both the control of the UAV and the detection of rescue targets, based on still images or video streams acquired by the UAV onboard cameras and sent to the ship using a wireless datalink. Achieving a bandwidth large enough to send visual feedback to the human operator with a proper sampling rate requires the use of dedicated hardware and can effectively decrease the search area, by limiting the UAV distance to the ship.
2) Alternatively, the UAV can follow an automated flight plan, using inertial and GNSS devices, while the target detection is still performed by the operator. This can reduce the number of operators required to work with several UAVs simultaneously, but a robust datalink for video transmission is still required: this reduces the UAV search area.
3) In the third and most advanced method the UAV follows an automated flight plan and performs onboard detection of castways, allowing the operator to only analyse the situations considered relevant by the automatic detector. By combining UAVs with automatic object detection algorithms we can create a system where the UAVs only send visual feedback (together with corresponding location) to the human operator when the onboard detector

has enough confidence on the presence of a rescue target in the acquired images. This method has two major advantages: on one way, a single human operator can potentially deal with several UAVs simultaneously deployed; on the other way, a large search area can be covered, since each UAV only sends images considered relevant by the detection algorithm and thus a larger distance from the ship can be achieved due to smaller bandwidth requirements.

This paper addresses the automatic detection of maritime rescue targets in maritime SAR missions, using standard RGB cameras on board the UAV. While the use of machine vision for detection of this kind of targets is not something new, there is not a vast literature on this topic. Sumimoto et al. [3] propose an image processing technique for the detection of rescue targets that resorts to color and shape information taken from aerial footage in order to detect possible rescue targets; however, using hand-picked features in the detection algorithm makes this algorithm prone to errors due to changes in environmental variables like *e.g.*, weather and light conditions, camera perspective, and distance from the UAV to the rescue target.

The work published by Mace in 2011 [4] tackled the issue of detecting marine debris, such as fishing nets, buoys and ropes. In this work, data from satellite observations and aircraft, in conjunction with mathematical models for the debris, was utilized to predict the location of such debris, but the author concluded that the detection of these debris remained a serious challenge to current technology at the time of publishing.

Ramirez et al. focused on sea rescue [5], proposing a coordinated sea rescue system based on an UAV and an Unmanned Surface Vessel (USV). In the system presented in that paper the USV benefits from the information provided by the UAV, which is capable of locating the castaways faster than the USV. The two subsystems were able to work in real-time with really simple defined behaviours. This system presents a good basis where the improvements of automatic Object Detection can be applied.

In 2017, Hoai and Phuong published a paper that studied the use of anomaly color detection on UAV images for SAR works. They were able to determine that in different SAR situations the appropriate color space and detection algorithm could be chosen to provide the best performance of anomaly detection [6].

In the same year, Dinnbier et al. published a paper that focused on using Gaussian Mixture Model (GMM) and Fourier Transforms for target detection in UAV maritime Search and Rescue [7]. Using GMMs they were able to remove background from images while leaving moving targets intact. By combining GMMs with Fourier Transforms they were able to improve their detection results.

In [8] a survey of traditional video processing for object detection and tracking in maritime environments is presented. However, in recent years, scientific advances in the field of deep neural networks have revolutionized several technological domains, from automatic image processing to speech processing and synthesis, including automatic text translation, among many other applications. In particular, the use of networks based on convolutional topologies brought a significant performance increase in the tasks of classification, detection and automatic image segmentation [9]. This kind of networks also began to be used, with great success, in maritime detection and segmentation tasks [10].

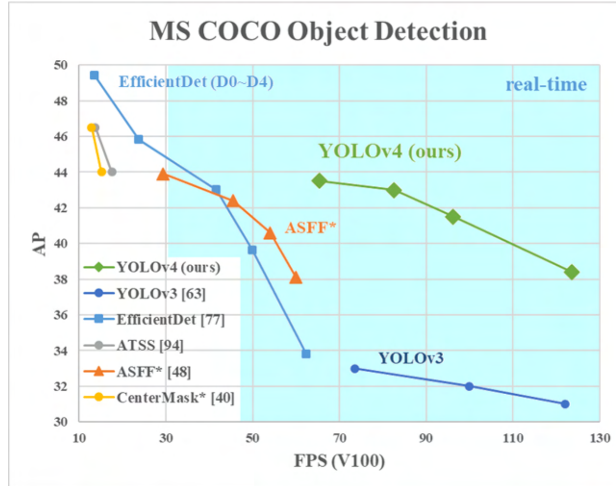

Fig. 2. Comparison of state-of-the-art detectors with respect to detection performance (AP) and speed (FPS) in the MS COCO dataset. Image taken from [11].

In particular, one-stage state-of-the-art detectors like YOLO v4 [11] currently achieve real-time inference capabilities while maintaining a very high detection accuracy (Fig. 2). Such advances make the application of such detectors to autonomous detection of rescue targets in SAR missions very promising: in this paper we present an automatic maritime rescue target detector, based on standard images acquired from an UAV standard RGB camera, based on recently developed deep learning detection techniques. No special hardware is needed to run such detection models, as a low-power embedded system like a Jetson Nano suffices to perform the detection in real-time.

The main contribution of this paper is the proof of concept that current state-of-the-art detection networks can be used in maritime SAR missions in a real-time manner, using simple embedded systems, with limited computational capabilities, on board the UAV. Additionally, a small dataset gathered and used for training the detection network is presented.

The remaining of the paper is organized as follows: Section 2 describes the data set acquired to train the detection network; Section 3 presents the neural network used and its training methodology; in Section 4 the results obtained are presented and in Section 5 conclusions are drawn and some possible future work is presented.

II. Data Acquisition

Due to its size, typically involving millions of parameters to be determined, the training of deep neural networks requires the use of massive amounts of data, in order to guarantee a correct convergence of these parameters, which allows to obtain a satisfactory performance in a test set that, following approximately the same statistical distribution as the data used for training the network, is nevertheless distinct from the same training set. Many detection networks are available in pre-trained versions, with parameters optimized for the detection of a certain number of predefined classes; Using these pre-trained neural networks is associated with two major advantages:

- You can immediately use this network when you want to detect objects corresponding to classes for which the network has been pre-trained;
- Even when the object to be detected does not belong to that set of classes of the neural network, it is generally much faster to carry out the readaptation of weights from the pre-trained network for this new class of objects, being necessary, for this purpose, the use of a much smaller set of data.

Bearing in mind that no current detection neural network is pre-trained to detect a "Rescue Target" class object, it became necessary to obtain and annotate a dataset that would allow the training of a network that considers this class. For this purpose, three different image sources were considered:

- Dataset 1: Images taken from videos on the YouTube site showing people at sea swimming and performing various aquatic activities (Fig. 3);
- Dataset 2: Images acquired from a DJI Mini 2 drone, close to a beach, of people in kayaks and simulating the castaway situation (Fig. 4);
- Dataset 3: Images captured from a Parrot Anafi drone of cadets from the Naval Academy in a field exercise in the Tagus river, during the swimming crossing, in the late afternoon, of a small channel (Fig. 5). In this situation, we tried to obtain images corresponding to different points of view, often with the sun in front of us to produce glare and reflections in the water.

The images of each dataset were manually annotated, in order to indicate the presence and location of each alleged rescue target in the image. In the end, a set of 550 annotated images was obtained for dataset 1, 67 for dataset 2 and 292 for dataset 3, in a total of 909 annotated images.

Considering that datasets 2 and 3 were more representative, due to the context in which they were acquired, of the situation corresponding to the presence of shipwrecked people in maritime images, it was decided

Fig. 3. Example images belonging to Dataset 1, taken from YouTube videos.

TABLE I

Summary table of the different datasets used in this work.

Training Datasets	# Images	Avg. number of objects
Dataset 1 (YouTube)	550	13
Dataset 2 (Beach)	53	2
Dataset 3 (River)	233	5
Testing Datasets	**# Images**	**Avg. number of objects**
Dataset 2 (Beach)	14	2
Dataset 3 (River)	59	5

to remove 20 % of annotated images from each of these datasets for a set independent test, ie respectively 14 and 59 images from datasets 2 and 3: these images were never used to train or adjust detection network parameters. The summary of the characteristics of these datasets is summarized in Table I.

III. Training

There are currently many detection networks for several classes of objects in RGB images: in general, for each of these networks, reducing the processing time required to process an image also leads to a decrease in the detection performance of the same network. Since in this work we intend to use a detection network capable of processing data in real time, even if it runs on hardware with reduced computational capacity, we focused on the YOLOv4-tiny network, as a consequence of its high real-time processing capacity while, at the same time, maintaining a very decent performance [11].

This YOLOv4-tiny network has a 38-layer structure and is reported to run at around 30fps on a Jetson Nano board, a small embedded computer specially designed for tasks related to artificial intelligence and vision and image processing, resorting to an optimized model using TensorRT [12]. A pre-trained version with the COCO dataset [13], with the capability to detect 80 different classes of objects, was used in this work to initialize the network weights, and then refined for a single class, corresponding to the rescue target, using the labelled aforementioned datasets.

Fig. 4. Example images belonging to Dataset 2, taken near a beach.

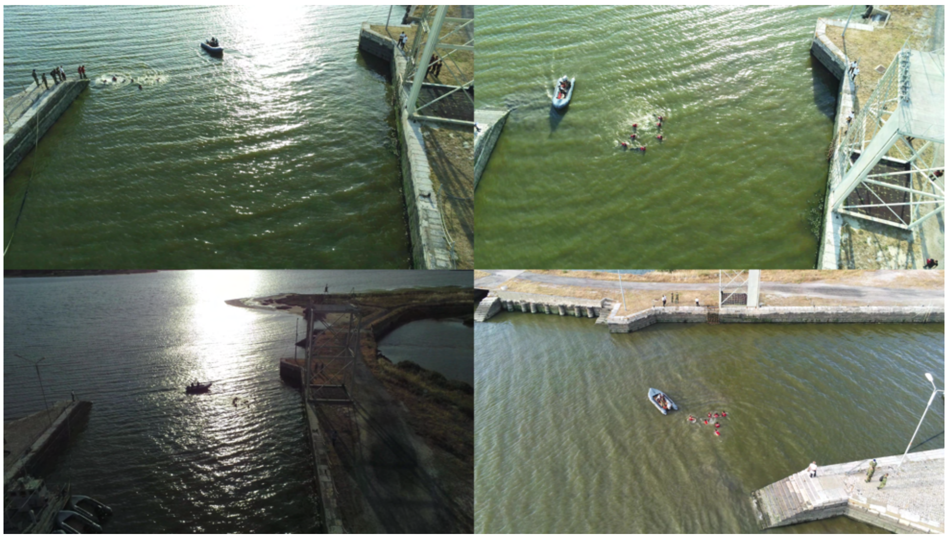

Fig. 5. Example images belonging to Dataset 3, obtained during the field exercises in the Portuguese Naval Academy.

In order to try to increase the performance of the network during the training for the detection of the new “Rescue Target” class, due to the reduced size of the training dataset, we proceeded to augment the existing data, through the application of several random transformations such as tonality change, image rotation and flipping, and adding random noise. The study of the influence of several network meta-parameters, such as the learning rate and the batch size, was also carried out, in order to try to obtain the best possible performance.

IV. Results

After training the YOLO network, using the set of images augmented from the 3 training datasets, a precision of 68 % and a recall of 55 % were obtained, for a detection threshold of 0.4, where precision and recall have the standard definitions:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

and

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} ,$$

where TP, FP and FN correspond to the number of detections, in the test set, categorized as true positives, false positives and false negatives, respectively. Each YOLO network detection in the image comes with a confidence value (between 0 and 1), and the detection threshold mentioned above corresponds to the minimum level of that confidence level for which a detection is considered, *i.e.*, detections with a confidence score less than 0.4 are discarded. By increasing the value of this threshold, there is more selectivity in what is considered a detection: this generally leads to an increase in precision and a corresponding decrease in recall. This value of 0.4 was chosen as a good compromise between detection precision and recall, based on the precision-recall curve obtained after training the YOLO network (Fig. 6). Since precision

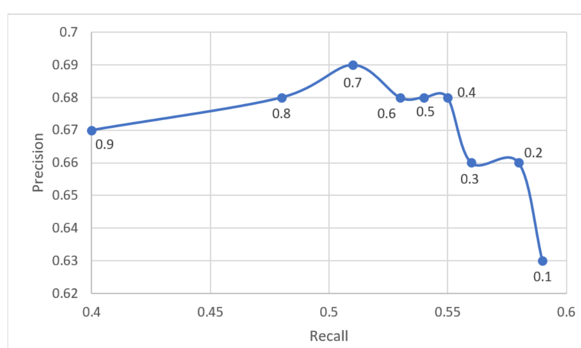

Fig. 6. Precision-Recall curve for several threshold values.

corresponds to the percentage of detections that are considered correct detections, and recall is the percentage of true positives in the dataset that are effectively detected by the network, this detection threshold can be changed to suit a particular objective. For instance, the threshold value can be lowered to increase the detection recall while lowering the precision: this increases the chance of detecting a victim in the water, at a cost of a higher rate of false detection positives reaching the human operator.

Fig. 7 compares the detection results of the YOLO network, before and after being trained with images from Dataset 1. Note that the pre-trained network does not have the "Rescue Target" class, but it is still able to detect objects of the "boat" and "surfboard" classes in the images. After training with this new class, using only dataset 1, the network can detect the presence of possible castaways in images from Dataset 2 in a satisfactory way.

The training sets used to train the network, as expected, decisively influence the final performance of the network and the results obtained: in Fig. 8, the detection in the scenario corresponding to Dataset 3, before and after training, is analysed.

Before training, using default YOLO weights and classes, the detector can only identify the boats in the two example images (Fig. 8, left). After training using Dataset 1 (YouTube images) the network already detects people in the water, while also erroneously identifying people on the vessel and on the dock as rescue targets (Fig. 8, middle). After training with the remaining datasets (2 and 3), the network is able to correct the false positives and false negatives that arose previously, as shown in Fig. 8, right.

Three small video sequences are available on YouTube to demonstrate de detection capabilities after the training process, corresponding evaluation on Dataset 1 (YouTube)[1], on Dataset 2 (Kayaks and Beach)[2] and on Dataset 3 (River)[3].

V. Conclusions and Future Work

The recent developments in detection using deep neural networks make it possible to easily integrate state-of-the-art detectors on board an UAV, without the need to use dedicated hardware, as some of these detectors are able to operate in real time in affordable and lightweight processing units for embedded applications, that are particularly suited for artificial intelligence and computer vision tasks, while still keeping a good detection accuracy. In this way, human operators no longer need to constantly monitor the video sequences acquired by an UAV performing a SAR mission, as the UAV will only transmit the most relevant frames, corresponding to a high confidence of having a rescue target. This reduces the bandwidth required in the UAV datalink and thus allows for an extended autonomy and search range.

Despite being carried out with a limited training set, we present, in this work, a proof of concept that automatic and robust detection of maritime rescue targets is possible if a diverse and representative dataset is gathered to train such a detector. Also, for practical applications, a filtering stage should be employed, resorting to temporal filtering techniques like Kalman or particle filters: this topic will be addressed in future work.

Acknowledgments

This work was supported by the Portuguese national project VOAMAIS (02/SAICT/2017/31172).

References

[1] "OECD: Ocean shipping and shipbuilding," https://www.oecd.org/ocean/topics/ocean-shipping/, accessed: February 2022.

[2] "International maritime organization: IAMSAR manual," https://www.imo.org/en/OurWork/Safety/Pages/IAMSARManual.aspx, accessed: February 2022.

[3] T. Sumimoto, K. Kuramoto, S. Okada, H. Miyauchi, M. Imade, H. Yamamoto, and T. Kunishi, "Machine vision for detection of the rescue target in the marine casualty," in *Proceedings of IECON'94-20th Annual Conference of IEEE Industrial Electronics*, vol. 2. IEEE, 1994, pp. 723–726.

[4] T. H. Mace, "At-sea detection of marine debris: overview of technologies, processes, issues, and options," *Marine Pollution Bulletin*, vol. 65, no. 1, pp. 23–27, 2012.

[5] F. F. Ramírez, D. S. Benitez, E. B. Portas, and J. A. L. Orozco, "Coordinated sea rescue system based on unmanned air vehicles and surface vessels," in *OCEANS 2011 IEEE-Spain*. IEEE, 2011, pp. 1–10.

[6] D. K. Hoai and N. Van Phuong, "Anomaly color detection on uav images for search and rescue works," in *2017 9th International Conference on Knowledge and Systems Engineering (KSE)*. IEEE, 2017, pp. 287–291.

[7] N. M. Dinnbier, Y. Thueux, A. Savvaris, and A. Tsourdos, "Target detection using gaussian mixture models and fourier transforms for uav maritime search and rescue," in *2017 International conference on unmanned aircraft systems (ICUAS)*. IEEE, 2017, pp. 1418–1424.

[8] 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," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 8, pp. 1993–2016, 2017.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[10] C. Pires, A. Bernardino, and B. Damas, "Ship segmentation in aerial images for maritime surveillance," in *26th Portuguese Conference on Pattern Recognition, RECPAD 2020*, 2020, pp. 81–82.

[11] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," https://arxiv.org/abs/2004.10934, 2020.

[12] "Examples demonstrating how to optimize Caffe/TensorFlow/DarkNet/PyTorch models with TensorRT," https://github.com/jkjung-avt/tensorrt_demos#yolov4, accessed: February 2022.

[13] "Microsoft Common Objects in Context (COCO) Dataset," https://cocodataset.org, accessed: February 2022.

[1] https://www.youtube.com/watch?v=sQuIbAba6CY
[2] https://www.youtube.com/watch?v=z33d4B3_f_I
[3] https://www.youtube.com/watch?v=8uQBeveMxkc

Fig. 7. Detecting rescue targets in Dataset 2: before training, using pre-trained YOLO weights (left) and after training with Dataset 1 (right).

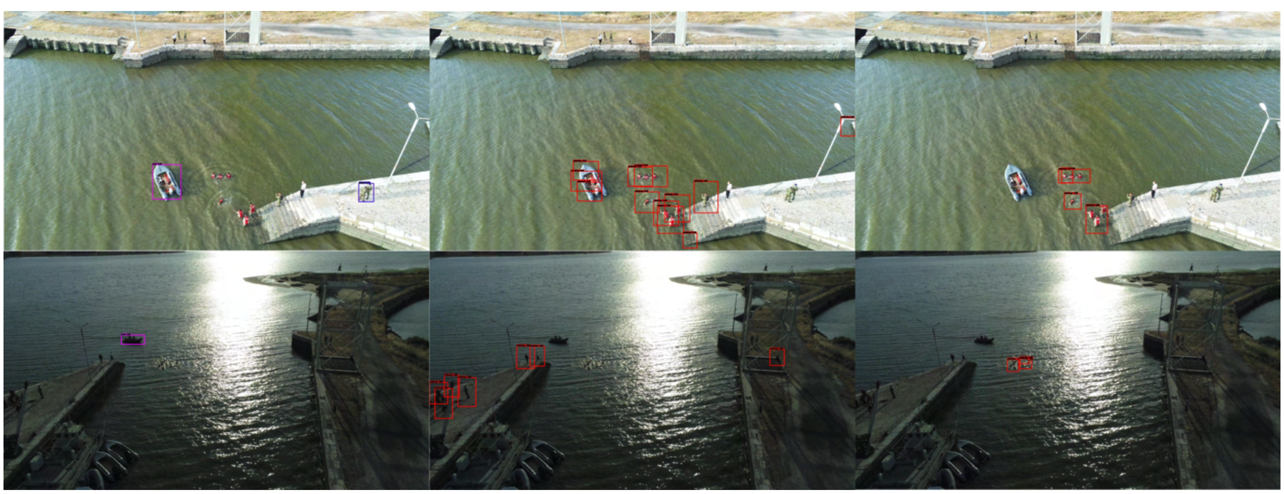

Fig. 8. Detecting rescue targets in Dataset 3: before training, using pre-trained YOLO weights (left), after training using Dataset 1 (middle) and after training with all Datasets (right).