Real-Time Automatic Wall Detection and Localization based on Side Scan Sonar Images

Martin Aubard[1], Ana Madureira[2], Luis Madureira[1], José Pinto[1]
[1]OceanScan-MST
[2]ISRC, ISEP/IPP
Porto, Portugal
{maubard,lmad,zepinto}@oceanscan-mst.com, amd@isep.ipp.pt

***Abstract*—Accurate identification of an uncertain underwater environment is one of the challenges of underwater robotics. Autonomous Underwater Vehicle (AUV) needs to understand its environment accurately to achieve autonomous tasks. The method proposed in this paper is a real-time automatic target recognition based on Side Scan Sonar images to detect and localize a harbor's wall. This paper explains real-time Side Scan Sonar image generation and compares three Deep Learning object detection algorithms (YOLOv5, YOLOv5-TR, and YOLOX) using transfer learning. The YOLOv5-TR algorithm has the most accurate detection with 99% during training, whereas the YOLOX provides the best accuracy of 91.3% for a recorded survey detection. The YOLOX algorithm realizes the flow chart validation's real-time detection and target localization.**

I. Introduction

The maritime economy and infrastructures have expanded in the last few years, which provides interest in maritime technology to protect and keep contributing to this area. Autonomous Underwater Vehicles (AUV) provide submarine information (features, infrastructures), making us aware of this environment. Considering the visible underwater uncertainty, the camera's utility is only used for specific tasks close to the bottom, or the target. The underwater environment is not well suited for optical images. These images are often dark and blurred, which does not provide great accuracy for object detection. However, acoustic sonar sensors are particularly well suited for underwater surveys [21]. They provide high range detection and are not influenced by the lack of luminosity. Side Scan Sonar (SSS) is the most common sonar used for seabed inspection and object detection. Through its two transducers localized on each part of the AUV, it scans a wide area and outcomes with a high-resolution sonar image of the seabed. Those images are mainly used to detect objects and features localized on the seabed. However, those images require high expertise in the maritime environment to detect and annotate the area of interest. With the improvement of Artificial Intelligence (AI) and specifically Deep Learning (DL) for object detection, several algorithms have started to be used in the last few years to process the data after the AUV maneuver. This technical improvement allows AUV users to be more independent in their data processing. Recent advances in pattern recognition and Machine Learning (ML) have greatly improved feature extraction from images. Driven by the impulse of the deep neural network, which gives breakthrough results for different AI challenges such as Natural Language Processing (NLP) [12], autonomous driving [8], or computational vision [14]. With the breakthrough result of AlexNet in 2012 [9], based on DL classification from images dataset, DL algorithm creation, and implementation have known a massive increase for computational vision fields such as classification [10], object detection [14] and image segmentation [2]. As in many fields, the DL algorithms for object detection gives excellent results for underwater images, especially sonar images. In this paper, we will compare the influence of 3 different DL algorithms for real-time object detection based on Side Scan Sonar images. Those algorithms are the YOLOv5 [18], the YOLOv5-TR [3], and the YOLOX [14]. This paper will compare the results of those three algorithms in real-time constraints for Side Scan Sonar implemented into an AUV. This paper is structured as follows: the section II introduces the current state of the art for real-time object detection, the section III explains the methodology used for real-time SSS images processing and real-time object detection, section IV explains in detail the sonar samples treatment and the sonar image generation, section V introduces the dataset created for this paper, the section VI explains the three algorithms (YOLOV5, YOLOV5-TR, YOLOX) training and result for offline experiments, and finally, the section VIII is the implementation of the algorithm for real-time detection using real-time processed sonar images.

II. State Of the Art

One of the main issue in object detection algorithm is to find the best balance between accuracy and efficiency using one or two stage detectors. R-CNN [11], Faster R-CNN [1], and Features Pyramid Network (FPN) are the current two-stages algorithms used for Side Scan Sonar images. Since 2018, and the advent of the one-stage detector YOLO (You Look Only Once) [15] algorithm, which was one the best balance between efficiency and accuracy; several updates of this algorithm has been proposed with YOLOv3, YOLOv4, YOLOv5. To overcome the shortcomings of the traditional manual detection of underwater targets in SSS images, a real-time automatic target recognition (ATR) method has been proposed in several papers [2], [4] and [1]. The paper [1] compares three different types of real-time ATR algorithms (the YOLOv3, YOLOv5, and the Faster RCNN) for pipeline inspection. The paper's outcome gives the same accuracy (97.62%) for the YOLOv5 and Faster RCNN models. However, the YOLOv5 algorithm is 62 times faster than the Faster RCNN, a tremendous advantage for a real-time utility. YOLOv5 is the fifth version of the YOLO family and is one of the most accurate and efficient in the States Of the Art (SOA) for object detection. The YOLOv5-TR [3] has the same architecture as the previous one; however, it is implemented with a Transformer module at the end of its backbone. Transformers have been a cutting-edge method in the NLP area, implementing a Multi-Head Attention module. "Attention" can be defined "as mapping query and a set of key-value pairs to an output, where the query, keys, values, and output are all the vectors"[6]. Implementing this method for NLP processing has been known as a significant improvement and outperformed the SOA. Due to its outperforming results, the transformer method has started to be implemented for computational vision by introducing the DEtection TRansformers (DETR) method for object detection [7], followed by DEit [16] and Vit[5]. In the YOLOv5-TR algorithm [3], the CNN YOLOv5 architecture is implemented by a transformer module with 4 Multi Head Self-Attention (MHSA) [6] between the backbone and the neck of the YOLOv5 algorithm. This algorithm is explained in detail in the section VI.

The YOLOX [14] algorithm is a new improvement of the YOLO series and tackles the current SOA in object detection. This algorithm uses YOLOv3 with the Darknet53 backbone and a Spatial Pyramid Pooling(SPP) layer as a baseline and uses Features Pyramid Network (FPN) and decoupled head for classification and regression. The YOLOX aims to implement the anchor-free model into the YOLO series.

III. Methodology

The methodology proposed and followed for this paper is depicted in Figure 1. This method shows the pipeline followed from the dataset creation to the implementation in real-time of the automatic target detection based on SSS.

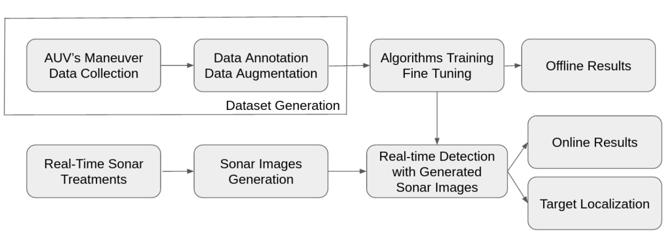

Figure 1: Pipeline Methodology

The implementation of DL algorithm requires a reliable and well-annotated dataset. However, it is complicated to find an open-source dataset in the underwater area. The lack of a dataset restrains the opportunity to train ML models with a massive amount of data. The first obligation that we have to do is to collect underwater data. For this task, we have used the Light Autonomous Underwater Vehicle (LAUV) [17] from Oceans-Scan MST company, set up with the SSS Klein3500. The maneuver took place at the Matosinhos harbor in Portugal, where different objects were detected. After the data collection, the data was displayed on the Neptus software [13], an AUV maneuver supervisor that allows replaying missions and displaying sensor information and visualization through its interface. Neptune provides crop tools to crop regions of interest or target objects into its replay data tool. This part aims to identify and separate targets and store them into a first dataset. Thanks to Neptune, this dataset can be extracted for the annotation process. The annotation process can require a maritime expert to identify the target accurately. For this paper, we are focusing on wall detection, which is a target that can be easily detected by non-expert maritime fields, as we will see in the section IV. Creating an annotated dataset allowed us to train the DL algorithms to use for this paper: the YOLOv5, YOLOv5-TR, and the YOLOX. The algorithm training and offline results are explained in detail in section IV. Regarding the results, the chapter ends with the approbation of one algorithm to implement in real-time object detection. This real-time object detection process requires real-time image generation allowing the DL algorithm to work in real-time. This process is focused on the real-time collection of SSS data and the treatments of the samples. Thanks to the SSS treatment, accurate and high qualities SSS images are generated. This process is detailed in the next chapter. The generation of real-time SSS images will allow the DL algorithm chosen in section VII to be tested in real-time. Dune [13], embedded software in charge of the control and sensors onboard the vehicle, will be used to test the real-time survey. The methodology proposed for Automatic Target Recognition is shown in Figure 2.

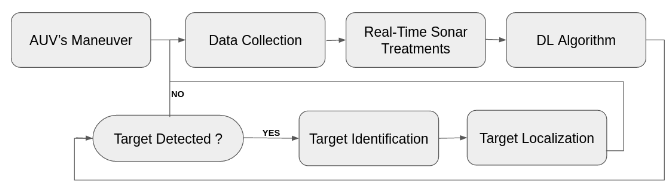

Figure 2: Implementation Methodology

This flow chart gives an overview of the real-time object detection and localization process, which is the outcome of this paper. This flow chart is divided into two main parts, which are the data collection with image generation and the second one is the object detection with target localization.

IV. Side Scan Sonar treatment

Side Scan Sonar sensor provides waterfall images updated by a continuous acoustic signal. This signal gives seabed intensities values and is usually quantified into 8, 32, or 64 bits ranges, meaning that each pixel value is quantified in one of those bit ranges. Pre-processing the SSS signal means re-quantified this range into an 8bits range value generating a 0-255 color map from the SSS value and generating comprehensible color images. The 0-255 range can be adjusted regarding the lowest and highest values recorded from the SSS. However, this pre-processing part is not enough to generate usable SSS images, as shown in Figure 3. Furthermore, the Klein 3500 sonar gives 2 different frequencies: 450KHz and 900KHz. For this paper we will be focus on the high frequency, which gives better accuracy for object detection.

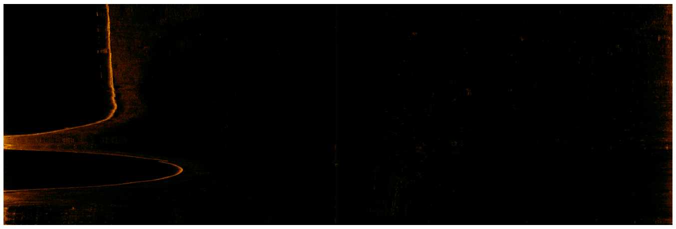

Figure 3: Side Scan Sonar Image

Some sonar signal treatments have to be realized. Acoustic waves spread and absorb losses when traveling through seawater. This loss is characterized by an information difference between the detection close to the transducer and the one far from it. The methodology used to correct and minimize this effect is called the Time Variable Gain (TVG), and the mathematics formula is given by equation 1:

$$TL = \alpha R + 20logR = \alpha(vt) + 20log(vt) \tag{1}$$

Where TL is the transmission loss, R is the spreading range expressed as propagation time t multiplied by sound velocity v, and α is the attenuation coefficient. The TVG result is normalized between 0 and 255, ensuring that the TVG normalization does not overpass the quantified SSS range.

However, even with the implementation of TVG and normalization, the generated images are not well suited for recognize object on it. Indeed, each pixel value has to be interpolated with its neighborhood value pixel, which harmonizes the final image result, as shown in Figure 4.

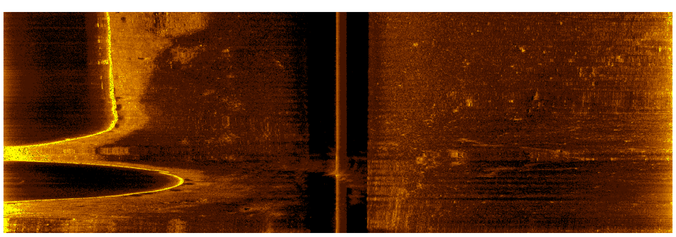

Figure 4: SSS Image with TVG, Normalization and Interpolation

This section will describe the algorithms used under this work focused on the generation of usable SSS images in real-time data collection. The data used for the last section of this paper provides from a survey done with a LAUV from Klein3500(SSS), and the real-time SSS values are replayed on the Dune embedded software. Three new lines are generated for each SSS ping, creating a new image that pushes out the last three lines of the images. This process is realized while the SSS is activated.

V. Dataset

The lack of an open-source underwater dataset is a real problem for training DL algorithms for underwater object detection. This constraint obliges DL researchers in the underwater field to do their own surveys to collect their data. For this paper, we have used the LAUV from OceanScan-MST company and the Klein3500 to collect SSS images. The survey aimed to collect SSS images of the harbor's walls. Those wall data will be implemented into the algorithms to realize wall detection and localization. The dataset has been built with two different classes: **wall** and **noWall**.

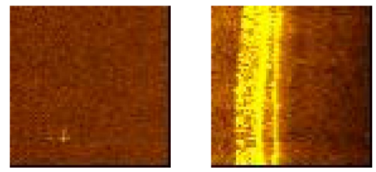

Figure 5: Dataset

The amount of **wall** and **noWall** data is shown on the Table 1.

Table 1: Dataset Information

Wall	noWall	Total
243	386	629

As we know, the amount of data is essential for a DL algorithm training process. DL algorithms are known to be data-driven and are affected by the amount of data. Thanks to several AUV maneuvers to collect **wall** and **noWall** data, those are only about 629 with 243 **walls** and 386 for **noWall**. To bypass the problem of the low amount of data, a technique called Data Augmentation is used to increase the dataset. Our dataset has been augmented using the clockwise and non-clockwise rotation, which increased the number of data. The final dataset is shown in table 2.

Table 2: Dataset Augmentation Information

Wall	noWall	Total
443	685	1128

To maintain a good distribution between the two classes, they have been mixed up and split into three groups: training, testing, and validation. The ratio used to split them is 83% for training, 11% for validation, and 6% for testing.

VI. Algorithms

This section describes in detail the three DL object detection algorithms use for SSS wall detection.

A. *YOLOv5 Algorithm*

YOLOv5 [18] algorithm is the fifth version of the YOLO family which stand for "You Look Only Once." YOLO is an object detection algorithm released in 2015 [15], has been adopted as a breakthrough result for real-time object detection giving a good balance between accuracy and efficiency. The algorithm using CNN structure, done as a regression problem, provides the class probabilities of the detected images and outcomes with bounding class probabilities. The whole process is realized in a single forward propagation. The YOLOv5's structure is shown in Figure 6.

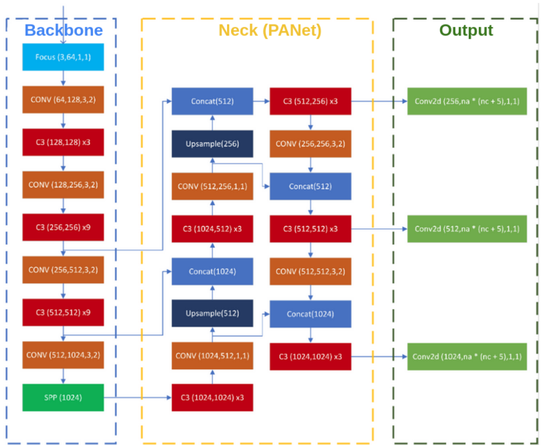

Figure 6: YOLOv5 Structure [18]

The model is split into three different parts: the backbone, the neck, and the output, which is called the model head. Model Backbone is mainly used to extract essential features from the input image. In YOLOv5, the Cross Stage Partial (CSP) Networks are used as a backbone to extract rich, informative features from an input image. The model neck is primarily used to generate models that generalize well in object scaling. The model Head is mainly used to perform the final detection part. It applies anchor boxes on features and generates final output vectors with class probabilities, objectness scores, and bounding boxes. In our case, the output will be bounding boxes with an accuracy percentage.

B. *YOLOv5-TR Algorithm*

Since the breakthrough results of the Transformers in the NLP area and the promising results of its implementation for computational vision through the End-to-End Object Detection with Transformers (DETR) [7], several CNN combined with Transformer module have been developed. The implementation of the Transformer module is localized at the end of the YOLOv5's Backbone, which is used to improve the quality of the input picture to improve the detection done by the Neck part. The Transformer module used for this architecture is shown in Figure 7. The Multi-Head Self Attention (MHSA) [6] aims to calculate the relation among pixels to make the algorithm more aware of the information on the picture. The implementation of the module helps to "take more Attention." Using a principle of query vector Q, key vector K, and value vector V, the MHSA used for this algorithm has four heads of Attention followed by a linear transformation.

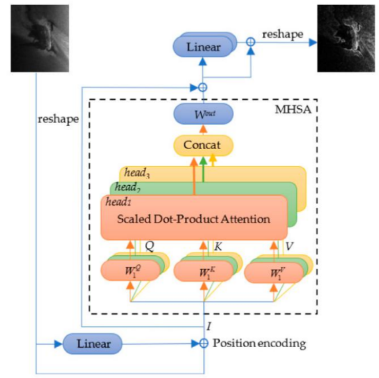

Figure 7: Multi-Head Self Attention [3]

C. *YOLOX Algorithm*

YOLOX algorithm has been proposed in 2021 [14]. This method took the YOLOv3-Darknet model as a baseline and improved it by implementing a decoupled head method separating the regression and annotation tasks.

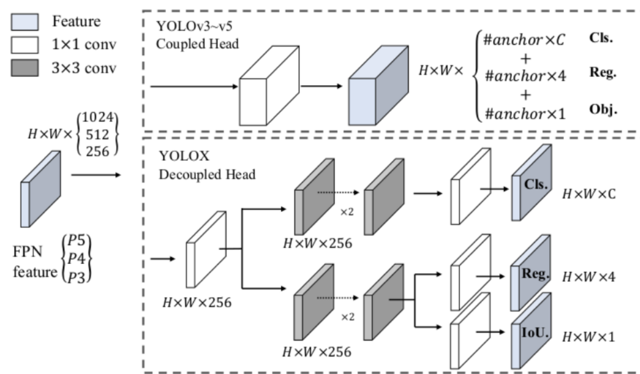

Figure 8: YOLOX Structure [14]

The Figure 8 shows YOLOX architecture, where for each FPN feature, a 1x1 Conv layer is used to reduce the feature channel to 256 and then add two different branches for classification and regression. The last significant advances in object detection are anchor-free detectors, advanced label assignment strategies, and end-to-end detectors. However, those improvements have not yet been implemented into the YOLO series, which this method is proposing. The single-stage object detection algorithms (e.g., YOLO) refine to the final detection location and are typically defined as the grid on the image coordinates at all possible locations, with different scales and aspect ratios.

VII. Training

This section focuses on the algorithms training and the comparison of the offline results. This part will outcome with the chosen algorithm for the real-time target detection tasks. First of all, all the algorithms have been implemented by pre-trained weights, which come from the well known COCO open-source dataset [20]. This method, called transfer learning, is a standard method used to pre-trained DL models with a massive amount of data which improves the training accuracy and efficiency of the model. This method can avoid the underwater dataset issue due to its low amount of data. The training has been processed with a single GPU Geforce RTX 3070 Ti for all the algorithms.

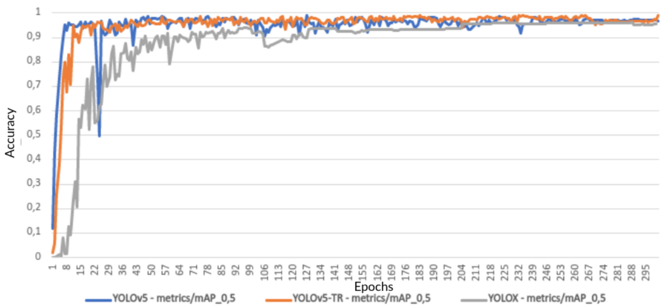

Figure 9: Training

The result shown in Figure 9 has been realized with a batch size of 12 for all the algorithms. The YOLOv5 and YOLOv5-TR algorithms have the best training behavior during the first epochs, which are already over 0.9% after 10 epochs. However, the YOLOX algorithm succeeds in reaching 0.9% at the 48th epoch and increases smoothly until the end of the training.

Table 3: Result Training

Algorithm	Best mAP_0.50 (%)	num Epoch
YOLOv5	98.9	128
YOLOv5_TR	99	266
YOLOX	96.2	218

The result shows that the YOLOv5-TR algorithm is the most accurate after training with 99%, followed by the YOLOv5 with 98.9% and the YOLOX with 96.2%. This training gives an advantage to the YOLOv5 algorithm implemented with the transformer module.

Regarding the training results, which are close to each other, an offline algorithm comparison is provided to get the best algorithm for object detection through SSS images.
The offline test consists of implementing SSS, recorded on Neptus and processed through a video object detection, with the algorithms trained. The SSS data has been taken in Porto's harbor but has not been used as training data, ensuring that the algorithms will not detect data that they have been trained with. The test compares the trained algorithm on a wall SSS data recorded offline. The comparison outcome will give the most suited algorithm for real-time wall detection. The data is about 14min of AUV survey, where 11.50min are data containing at least one wall. Table 4 shows the result.

Table 4: SSS offline validation

Algorithm	Wall Detected(min)	Total	FPS
YOLOv5	6.32	54.95%	90
YOLOv5_TR	10	86.95%	133
YOLOX	10.50	91.3%	125

The first version of the YOLOv5 algorithm is outperformed by its transformer improvement and the YOLOX, a YOLO3 improvement based on a free anchor. The most accurate algorithm for SSS detection is the YOLOX with 91.3% wall detected, which is almost 5% more accurate than the YOLOV5-TR. During the whole detection process, on average, the bounding box square was closer to the target with the YOLOX, which gives a better IOU value of the ground truth with a better prediction score. Regarding the frame per second (FPS) parameter, the YOLOv5-TR is the most efficient, with 133 FPS, 8 more than YOLOX, and 43 better than YOLOv5. An average AUV velocity during a survey is about a few meters per second, and the real-time SSS processing is relatively slow, which does not require a high FPS value to make the algorithm detect walls. However, the FPS acquired changes depend on the software and hardware used for the detection, which should be considered when the algorithm is used on an embedded vehicle. Thanks to this result, the algorithm chosen for the real-time wall detection based on SSS is the YOLOX.

VIII. Results

The section aims to use the algorithm chosen in the last section for real-time detection and localization of the target. This task would be helpful for several tasks, such as target/features detection on the seabed. In this paper, the algorithms have been trained for wall detection. This task is processed using Dune and Neptus software to create a real-time environment. Neptus replays a mission, and Dune receives the AUV information through the IMC protocol. Thanks to the SSS Treatment section, images are generated while the mission is replayed, and the algorithm processes each image to detect a wall. A bounding box is displayed on the image for each wall detected. Regarding the SSS characteristic, the information is about 75 meters on each side of the AUV. However, several parameters of the images have to be filtered to make the detection more efficient. First, below the AUV, a region called Nadir Gap is always displayed on SSS images. It refers to the lack of information region between the two transducers, and its size varies according to the distance from the AUV to the seabed. This region is displayed as a high luminosity straight line between two black straight lines. This representation on SSS images may disturb the algorithm because of its similarity with a wall. The Nadir Gap region is filtered from the image to reduce the algorithm disturbance. During the actual time of SSS image generation, some noise has been noticed for each transducer at the end of their range which is about 1 meter per each, and has also been filtered. Regarding those two image modifications, the range detection is now over 145.5 meters reducing the range detection to 4.5 meters. This new parameter is primordial to localizing the wall on the image, reducing the number of pixels over each line from 1563 to 1518. Each range meter is characterized by 10.42 pixels. Concerning these values, localizing a wall on the image from the AUV is possible. Thus, while the AUV maneuvered, the robot received wall position messages containing the target's side (port/starboard) and its distance from the AUV, as shown in Figure 10.

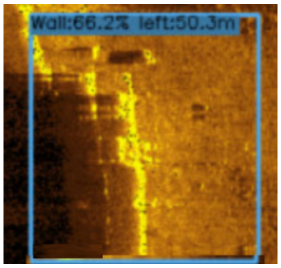

Figure 10: Wall Detection and Localization

IX. Conclusion

This paper aims to detect and localize walls in real-time through SSS images. It clusters several essential topics: the real-time SSS image generation, the most suitable Object Detection algorithm, and their implementation into a real-time AUV survey scenario. Since its release in 2020, the YOLOv5 has known several improvements, and the YOLOv5-TR is one of its most successful. This improvement is characterized by implementing a transformer module between its backbone and its neck. This module attempts to improve the accuracy of the detection by improving the image quality. Its training accuracy is about 99%, and its detection percentage on the offline validation is about 86.9%, which outperforms the first version of the YOLOv5 by 32%. In 2021, the YOLOX algorithm was released and proclaimed a better performance than the YOLO family based on a free anchor detection. This paper compared it with the YOLOv5 and its transformer improvement, the YOLOv5-TR. Its best mAP_0.50 is the lowest with 96.2%. However, it gives the best accuracy during the offline validation with a 91.3% of wall detected. This algorithm has been chosen for the real-time wall detection based on SSS images realized through Dune and Neptus software to simulate a real-time AUV survey. Thanks to the YOLOX bounding box, the outcome is the localization in real-time of the walls detected during the maneuver. Several future works can be conducted to improve this paper. As seen in the algorithm comparison, transformer modules improve object detection accuracy. For future works, this module could be implemented into the YOLOX algorithm [19], which could be an improvement for small object detection for underwater SSS images. This method can be applied for pipeline following regarding the proposed wall detection and localization. Another work should focus on the detected target's GPS absolute localization coordinate (long/lat). Indeed, for the moment, the localization message exchanges the position side and its distance from the AUV, which is accurate for a wall localization. However, better localization accuracy should be provided for a different target (e.g., mine). This localization improvement would give a better understanding of the AUV's environment.

Acknowledgements : This project has received funding from the European Union's EU Framework Programme for Research and Innovation Horizon 2020 under the Grant Agreement No 956200.

References

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," arXiv:1506.01497 [cs], Jan. 2016, Accessed: Feb. 09, 2022. [Online]. Available: http://arxiv.org/abs/1506.01497

[2] K. Li et al., "Real-Time Segmentation of Side Scan Sonar Imagery for AUVs," p. 5.

[3] Y. Yu, J. Zhao, Q. Gong, C. Huang, G. Zheng, and J. Ma, "Real-Time Underwater Maritime Object Detection in Side-Scan Sonar Images Based on Transformer-YOLOv5," Remote Sensing, vol. 13, no. 18, p. 3555, Sep. 2021, doi: 10.3390/rs13183555.

[4] J. Yang, J. P. Wilson, and S. Gupta, "DARE: AI-based Diver Action Recognition System using Multi-Channel CNNs for AUV Supervision," arXiv:2011.07713 [cs], Nov. 2020, Accessed: Feb. 09, 2022. [Online]. Available: http://arxiv.org/abs/2011.07713

[5] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv, Jun. 03, 2021. Accessed: May 17, 2022. [Online]. Available: http://arxiv.org/abs/2010.11929

[6] A. Vaswani et al., "Attention Is All You Need." arXiv, Dec. 05, 2017. Accessed: May 17, 2022. [Online]. Available: http://arxiv.org/abs/1706.03762

[7] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End Object Detection with Transformers." arXiv, May 28, 2020. Accessed: May 17, 2022. [Online]. Available: http://arxiv.org/abs/2005.12872

[8] S. Atakishiyev, M. Salameh, H. Yao, and R. Goebel, "Explainable artificial intelligence for autonomous driving: An overview and guide for future research directions." arXiv, Apr. 27, 2022. Accessed: May 17, 2022. [Online]. Available: http://arxiv.org/abs/2112.11561

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[10] X. Qin, X. Luo, Z. Wu, and J. Shang, "Optimizing the Sediment Classification of Small Side-Scan Sonar Images Based on Deep Learning," IEEE Access, vol. 9, pp. 29416–29428, 2021, doi: 10.1109/ACCESS.2021.3052206.

[11] R. Girshick et al., "Rich feature hierarchies for accurate object detection and semantic segmentation." arXiv, Oct. 22, 2014. Accessed: May 17, 2022. [Online]. Available: http://arxiv.org/abs/1311.2524

[12] K. Knight and D. Marcu, "Summarization beyond sentence extraction: A probabilistic approach to sentence compression," Artificial Intelligence, vol. 139, no. 1, pp. 91–107, Jul. 2002, doi: 10.1016/S0004-3702(02)00222-9.

[13] J. Pinto et al., "The LSTS toolchain for networked vehicle systems," in 2013 MTS/IEEE OCEANS - Bergen, Bergen, Jun. 2013, pp. 1–9. doi: 10.1109/OCEANS-Bergen.2013.6608148.

[14] Z. Ge et al., "YOLOX: Exceeding YOLO Series in 2021." arXiv, Aug. 05, 2021. Accessed: May 17, 2022. [Online]. Available: http://arxiv.org/abs/2107.08430

[15] J. Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," arXiv:1506.02640 [cs], May 2016, Accessed: Mar. 29, 2022. [Online]. Available: http://arxiv.org/abs/1506.02640

[16] H. Touvron et al., "Training data-efficient image transformers distillation through attention." arXiv, Jan. 15, 2021. Accessed: Jul. 26, 2022. [Online]. Available: http://arxiv.org/abs/2012.12877

[17] A. Sousa et al., "LAUV: The Man-Portable Autonomous Underwater Vehicle," IFAC Proceedings Volumes, vol. 45, no. 5, pp. 268–274, 2012, doi: 10.3182/20120410-3-PT-4028.00045.

[18] Ultralytics-Yolov5. Available online: https://github.com/ultralytics/yolov5 (accessed on 1 January 2021)

[19] J. Zhang and S. Ke, "Improved YOLOX Fire Scenario Detection Method," Wireless Communications and Mobile Computing, vol. 2022, pp. 1–8, Mar. 2022, doi: 10.1155/2022/9666265.

[20] T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context." arXiv, Feb. 20, 2015. Accessed: May 30, 2022. [Online]. Available: http://arxiv.org/abs/1405.0312

[21] H. Yu et al., "Bottom Detection Method of Side-Scan Sonar Image for AUV Missions," Complexity, vol. 2020, pp. 1–9, Oct. 2020, doi: 10.1155/2020/8890410.