Deep Learning for on-board AUV Automatic Target Recognition for Optical and Acoustic imagery

Leonardo Zacchini*,**, Alessandro Ridolfi*,**, Alberto Topini*,**, Nicola Secciani*,**, Alessandro Bucci*,**, Edoardo Topini*,**, and Benedetto Allotta*,**

*Department of Industrial Engineering, University of Florence, via di Santa Marta 3, 50139, Florence, Italy (e-mail: leonardo.zacchini@unifi.it, a.ridolfi@unifi.it).

**Interuniversity Center of Integrated Systems for the Marine Environment (ISME), www.isme.unige.it

***Abstract*—In the widespread field of underwater robotics applications, the demand for increasingly intelligent vehicles is leading to the development of Autonomous Underwater Vehicles (AUVs) with the capability of understanding and engaging the surrounding environment. Consequently, the automatic recognition of targets is becoming one of the most investigated topics and Deep Learning-based strategies have shown astonishing results. In the context of this work, two different neural network architectures, based on the Single Shot Multibox Detector (SSD) and on the Faster Region-based Convolutional Neural Network (Faster R-CNN), have been trained and validated, respectively, on optical and acoustic datasets. The models have been trained with the images acquired by FeelHippo AUV during the European Robotics League (ERL) competition, which took place in La Spezia, Italy, in July 2018. The proposed ATR strategy has then been validated with FeelHippo AUV in an on-board post-processing stage by exploiting the images provided by both a 2D Forward Looking Sonar (FLS) as well as an IP camera mounted on-board on the vehicle.**

***Index Terms*—Marine Robotics, Artificial Intelligence, Automatic Target Recognition, Autonomous Underwater Vehicles, Neural Networks,Machine learning for environmental applications.**

I. Introduction

Over the last few years, Deep Learning (DL) techniques have achieved significant success in digital image processing by fulfilling the object detection and classification tasks in increasingly challenging environments and scenarios. As a consequence, DL has recently resulted as the state-of-the-art approach in performing Automatic Target Recognition (ATR) by means of highly nonlinear feature extraction. In particular, modern Deep Neural Networks, such as the Faster R-CNN [1] and the Single Shot Multibox Detector (SSD) [2], have shown the potential to address the automated object recognition task in the underwater environment.

Autonomous Underwater Vehicles (AUVs) are commonly equipped with several payload sensors, including cameras and sonars, with the aim of perceiving and inspecting the subsea environment. Although modern cameras provide high-resolution images, optical data have the non-negligible drawback to significantly degrade in the presence of turbid water and low-light conditions. Conversely, acoustic sensors, such as Forward-Looking Sonar (FLS) or Side Scan Sonar (SSS), supply lower resolution, high-noise images with a wide range of coverage. As a result of the highlighted patterns, several studies have been proposed in order to extend the traditional object recognition DL techniques to underwater scenarios by exploiting acoustic images [3].

In the context of this work, a DL-based ATR architecture has been designed and implemented by using camera as well as sonar frames. A large image dataset has been collected, pre-processed and labeled in order to train the SSD model and Faster R-CNN for optical and acoustic images, respectively. Afterward, the trained neural networks have been incorporated in a custom ATR software, developed in the Robot Operating System framework. All the presented results have been validated by sea trials conducted with FeelHippo AUV, developed by Mechatronics and Dynamic Modeling Laboratory (MDM Lab) [4] of the Department of Industrial Engineering of the University of Florence (UNIFI DIEF).

II. On-board AUV ATR strategy

A. DL model selection

Since carrying out ATR while AUV navigating was the major purpose of the project, the focus has shifted to the SSD and Faster R-CNN architectures which guarantee the required trade-off between high-standard inference performance and the feasibility for real-time implementation. More in detail, the SSD network has been selected to fulfill a high-FPS recognition task with optical images. On the other hand, since the acoustic frames were captured with a lower frame-rate (3 Hz), the *mean Average Precision* (mAP) has been favored as model selection metric over the inference speed; as a consequence, Faster R-CNN has been preferred to faster but less accurate DL structures. Since the process of gathering a large dataset in an underwater scenario is by no means straightforward, exploiting transfer learning, by fine-tuning higher-order feature representations, allows to remarkably speed up the training phase. Thus, using pre-trained model weights has resulted as the optimal solution in terms of learning and convergence timings. As far as the specific selections are concerned, the SSDMobileNet v2 [2] and the Faster R-CNN Inception v2 [1] networks have been adopted to process, respectively, the optical and acoustic images.

B. Deep Neural Network Training

The training dataset was acquired with FeelHippo AUV during sea trials at the European Robotics League Emergency 2018 [5], which took place in La Spezia (Italy). The ERL Challenge was structured to address a simulated yacht accident in the basin of the NATO Science and Technology Organization Centre for Maritime Research and Experimentation (CMRE); as a partial task of a more complex mission, the AUVs were required to automatically recognize a damaged pipeline (represented by a yellow pipeline with attached a red marker) alongside a whole pipeline structure assembled by the aforementioned damaged pipeline as well as other five pipelines. To this end, 500 optical images of the pipelines and the red marker, with a resolution of 704×576 pixels, were acquired with the downward-pointing IPCam ELP 720p to train the SSDMobileNet v2 model. While 200 acoustic images in a native resolution of 894×477 pixels, depicting the structure, provided by the 2D FLS Teledyne Blueview M900, were used to train the Faster R-CNN Inception v2 network.

C. On-field validation

To validate the here proposed ATR strategy, a hierarchical-stage strategy has been employed. Firstly, several optical frames and acoustic images have been acquired by using, respectively, the bottom-looking ELP 720p MINI IP camera and the Teledyne BlueVIew M900 2D FLS during a FeelHippo AUV pre-programmed mission. In a second post-processing stage, the SSD and Faster R-CNN trained models have been executed on different dedicated hardware platforms; indeed, this hardware-decoupling solution provides the ATR system with the capability to process the images with the requested FPS value and guarantees the trained CNN modes to be real-time on-board runnable. With regard to the optical ATR approach, the SSD trained architecture has been optimized as a compiled graph, which has been subsequently loaded onto the Intel Neural Compute Stick 2, where it can run at 5.0 fps. Turning to the Faster R-CNN trained network, the prediction task on the acoustic frames have been fulfilled by means of the NVIDIA Jetson Nano, which managed to run the network at 1.0 fps. The developed ATR framework for AUV on-board applications is summarized in Fig. 1.

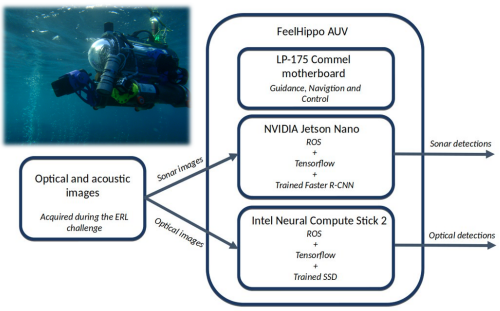

Fig. 1: The developed ATR strategy used to analyze collected images with FeelHippo AUV (top-left) during a pre-planned mission.

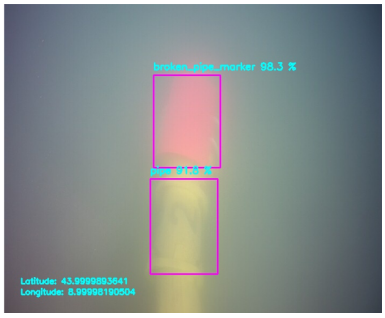

Fig. 2: Example of pipe and marker recognition in an optical image.

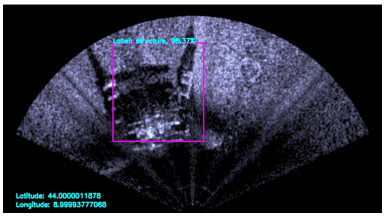

Fig. 3: Example of structure recognition in a 2D FLS acoustic image.

III. Conclusions and Future Works

The proposed work arises as a validation proof of the feasibility of DL methodologies for ATR tasks in the underwater environment on both the acoustic and optical subsea imagery. An experimental dataset has been acquired by using the payload sensors of FeelHippo AUV so as to optimally maximize the training stage effectiveness. As far as the DNNs, used in the training stages, are concerned, whilst a SSD network has been trained for the ATR task in optical images, a Faster R-CNN has been employed to develop an accurate trained model for the FLS acoustic imagery. Experimental tests have been carried out in order to validate the above-mentioned trained models loaded on dedicated hardware platforms. Different CNN architectures to fit the trade-off between inference speed and accuracy will be evaluated. Future works will also include the employment of the proposed ATR strategy in an overall intelligent system which led the vehicle to detect and recognize unknown targets as well as navigate towards them.

References

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, pp. 91–99, 2015.

[2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *European conference on computer vision*, pp. 21–37, 2016.

[3] M. Valdenegro-Toro, "Object recognition in forward-looking sonar images with Convolutional Neural Networks," in *OCEANS 2016 MTS/IEEE Monterey*, pp. 1–6, 2016.

[4] B. Allotta, R. Conti, R. Costanzi, F. Fanelli, J. Gelli, E. Meli, N. Monni, A. Ridolfi, and A. Rindi, "A low cost autonomous underwater vehicle for patrolling and monitoring," *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, vol. 231, no. 3, p. 740–749, 2017.

[5] G. Ferri, F. Ferreira, and V. Djapic, "Multi-domain robotics competitions: The CMRE experience from SAUC-E to the European Robotics League Emergency Robots," in *OCEANS 2017-Aberdeen*, pp. 1–7, 2017.