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.