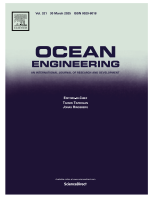
Ocean Engineering 340 (2025) 122341

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng



Research paper

Human detection in marine disaster search and rescue scenario: a multi-modal early fusion approach

Filippo Ponzini [a,*], David Van Hamme [b], Michele Martelli [a]

[a] Department of Marine, Electrical, Electronics and Telecommunication Engineering and Naval Architecture (DITEN), Polytechnic School, University of Genoa, Via Montallegro 1, Genoa, 16145, Italy

[b] Department of Telecommunications and Information Processing (TELIN), Faculty of Engineering and Architecture, Ghent University IPI-imec, Sint-Pietersnieuwstraat 41, Ghent, 9000, Belgium

ARTICLE INFO

Keywords:
Thermal camera
LiDAR
Search and rescue
Situational awareness
Marine autonomous surface ship

ABSTRACT

Marine disasters pose significant risks to both victims and rescuers, often occurring in challenging conditions. To address this, we propose a robust perception system for human-in-water search and rescue, leveraging early fusion of thermal imaging and LiDAR data. The system employs the YOLOv8 deep neural network to detect and classify survivors from multi-source images, maintaining high reliability even in adverse environments. The framework has been designed, implemented, and evaluated on a newly collected real-world dataset, specifically created for this application. Additional data augmentation simulates harsh operational and environmental conditions to enhance robustness. Performance was assessed in terms of detection accuracy and computational efficiency. The proposed multi-sensor approach achieved a precision of 93.5 % and a recall of 94.2 %, outperforming single-sensor models and demonstrating superior generalization in complex scenarios. Additionally, it reduces computational cost by approximately 64 % compared to a late fusion strategy, supporting efficient real-time processing. These results confirm that the proposed system significantly improves perception capabilities while meeting real-time constraints, making it suitable for deployment in time-critical maritime rescue operations. By integrating autonomous sensing and intelligent processing, this work contributes to safer and more effective search and rescue missions at sea.

1. Introduction

Maritime Search and Rescue (SAR) operations are crucial for saving lives at sea, often involving complex missions under challenging conditions. Traditionally carried out with human-crewed ships and aircraft, these efforts can be time-consuming, costly, and risky for rescuers. With modern technological advancements, autonomous vehicles are becoming increasingly important in a wide range of applications. Despite the successful deployment of Unmanned Aerial Vehicles (UAVs), including in SAR operations across many domains, their use in maritime SAR operations is less effective due to their limited range, high vulnerability to weather conditions, and inability to provide direct intervention. In this context, the use of a Unmanned Surface Vehicle (USV) or Autonomous Surface Vehicle (ASV) is preferable, as it is inherently more resistant to atmospheric events, can carry larger payloads of essential relief supplies (such as food, water, and medicine), and can transport flotation aids or inflatable rafts, as shown by Matos et al. (2013), that can be easily deployed and towed by the vehicle itself. The use of ASVs/USVs for challenging missions is a trend supported by the work of numerous research groups, spanning applications from the surveillance of sensitive and high-risk areas (Ponzini et al., 2024), environmental data collection (Odetti et al., 2024), and SAR missions (Akbar et al., 2022; Mansor et al., 2021), to support for warfare operations (Boretti, 2024).

In SAR missions, uncrewed vessels are appreciated as they can cover large areas quickly, operate in dangerous environments, and work in coordinated swarms to improve search efficiency; moreover, they can be deployed directly by aerial vehicles in the disaster area. The use of ASVs represents a major step forward in making SAR operations faster, safer, and more effective, helping to protect lives while optimizing resources.

This article proposes the development of a perception system to identify survivors adrift at sea during a SAR operation. This capability is fundamental and serves as an enabling technology for SAR operations with Autonomous Surface Vehicles. Furthermore, it represents a crucial aid in Decision Support Systems (DSS) for USV remote operators or personnel directly involved in the search onboard manned vessels. Specific attention is given to the system's robustness against adverse weather

* Corresponding author.
E-mail address: filippo.ponzini@edu.unige.it (F. Ponzini).

https://doi.org/10.1016/j.oceaneng.2025.122341
Received 5 June 2025; Received in revised form 21 July 2025; Accepted 28 July 2025
Available online 3 August 2025

and sea conditions, as these are key contributing factors to maritime accidents.

The paper is structured as follows. Section 2 gives an overview of trends and contributing factors in maritime accidents and SAR operations. In Section 3, the related work is presented. In Section 4, the methodology and materials used are shown. In Section 5, the results obtained are presented and discussed, with analysis provided in Section 6. Finally, in Section 7, conclusions are drawn and future developments are explained.

2. Background

To formulate choices and methodologies adequately for such a system, it is essential first to define the characteristics of typical SAR scenarios. Therefore, an introductory investigation into the most frequent circumstances requiring the initiation of a SAR mission has been conducted, building upon the understanding of maritime accident dynamics. The EMSA Annual Overview of Marine Casualties and Incidents 2024 (, EMSA) analyzes accident event types from 2014 to 2023, categorizing them as 'Human action', 'System/equipment failure', 'Hazardous material', and 'Unknown'. The analysis reveals that "Environment" is the primary and most significant contributing factor in three out of the four accident event types, and it is the second most crucial factor in the remaining category, "Human action".

Each year, the US Coast Guard (USCG) publishes a report detailing the dynamics and causes of recreational boating accidents. The 2023 report (Guard, 2023) reveals several key trends. Despite a higher overall number of recreational boating accidents during the summer months (due to the higher recreational traffic), the percentage of fatal accidents is higher in autumn and winter, with adverse weather conditions appearing to be a crucial factor (e.g., 29 % fatality rate in December, 22 % in October, compared to a 14 % average). Similarly, while the highest volume of accidents occurs during daylight hours with heavier traffic, the fatality rate is considerably higher at night when visibility is poor (19-23 % between 10 PM and 2:30 AM versus a 14 % daily average). Directly attributable environmental conditions were the primary cause of 558 accidents out of 3844, resulting in 100 deaths and 257 injuries; an additional 44 accidents were related to restricted vision. While the majority of accidents occur in good visibility due to increased boating activity, accidents in low visibility, although less frequent, exhibit a substantially higher injury and fatality rate (82 % against 68 %). Finally, drowning remains the leading cause of death in recreational boating accidents, accounting for 377 of the 564 fatalities in 2023.

Liu et al. (2021) provide a systematic analysis for maritime accident causation in Chinese coastal waters using machine learning approaches; the study concludes that bad weather conditions, such as fog, rain, and rough seas, often lead to catastrophic accidents. Additionally, the authors note that accident probability is higher at night than during the day, and fog is confirmed as a critical weather condition for marine accidents.

Brandt et al. (2024) analyze machine learning models for accident risk prediction by integrating weather data obtained from 1982 to 2021 by the Norwegian Maritime Authorities (NMA). The study revealed that the leading weather variables for accident prediction are visibility, wind, sea level pressure, and moon phase.

Panagiotidis et al. (2021) provide a comprehensive review of existing marine accident datasets, confirming that adverse weather conditions are strongly related to fatalities. In particular, wind and fog intensity seem to be proportional to the ratio of deaths to injuries.

Zhang et al. (2019) confirm the link between adverse weather conditions and low visibility with the severity of the accident and fatality ratio.

Events that support the thesis of poor visibility in an SAR scenario can be found in past disasters. In 1955, the Shiun Maru disaster resulted in 168 fatalities in thick fog. In 1956, SS Andrea Doria and MS Stockholm collided during the night (around 11:00 p.m.) in thick fog, resulting in 46 deaths. The Moby Prince disaster, which cost the lives of 140 people off the coast of Livorno (Italy) in 1991, occurred at night (around 10.00 p.m.) and during the legal process (Criminal Court of Livorno, Section I, 1998), the presence of advection fog emerged as a contributory cause of the dramatic collision. In addition, the ship was engulfed in flames, as was the surrounding sea, creating a thick blanket of smoke that made rescue efforts even more difficult. In the recent North Sea collision between container ship MV Solong and oil tanker MV Stena Immaculate (occurred in March 2025), heavy fog was registered by onboard camera; moreover, the ships burned, producing a thick blanket of smoke in the following hours (Bryony Gooch, 2025).

Given the insights into maritime accidents and SAR challenges, a likely SAR scenario for survivor search involves adverse weather and sea states, characterized by green seas, a high probability of nighttime conditions, and the potential for dense fog. Furthermore, the presence of wreckage and debris, along with the possibility of burnt objects and thick smoke, cannot be excluded. Considering such a challenging scenario, the aim is to develop a robust perceptive system for quick survivor detection, specifically designed to equip an ASV and enable the successful execution of the SAR mission.

3. Related work

This section reviews studies focused on human detection in maritime Search and Rescue (SAR) and Man Over Board (MOB) scenarios. While many works address object detection at sea, this review focuses specifically on human detection in SAR contexts. In Gennarelli et al. (2022), a feasibility study of floating life-jacket detection using an FMCW MIMO Radar is provided.

Mansor et al. (2021), Jian et al. (2017) propose a sonar-based ASV platform for SAR operations. In Akbar et al. (2022), a SAR-oriented ASV platform equipped with an RGB camera is proposed. The detection is entrusted to image-only and is based on a YOLO convolutional neural network (CNN) detector.

In Kang et al. (2020), a complete ASV platform for the SAR mission is presented. A drowning detection module is introduced; the detection is carried out using a non-specified deep learning neural network trained on Unity-generated images. In Taipalmaa et al. (2024), Wang et al. (2023), Ancy Micheal and Sivaramakrishnan (2024), Lygouras et al. (2019), Rizk et al. (2023) a UAV is equipped with a visible-light optical sensor for human-in-water detection; the procedure is based on an image-only CNN, tested in calm water and good visibility scenarios.

Li et al. (2021) show RGB camera human-in-water detection based on a YOLO CNN. In Cheong et al. (2024), a thermal camera-based method is presented for human-in-the-water detection, tested on real acquired and simulated data. The procedure is based on unsupervised Domain Adaptation followed by a segmentation step. Feraru et al. (2020) propose thermal camera-equipped UAV for person-in-water detection using Faster R–CNN. Martins et al. (2013) propose the use of RGB and thermal cameras onboard an ASV for detecting humans in water. However, the data from the two sensors is not fused; instead, the processing relies on histogram-based thresholding techniques. In Katsamenis et al. (2020), various approaches for man overboard detection are analyzed, including the use of the YOLO framework on both thermal and RGB images. However, no sensor fusion techniques are explored. The study emphasizes the importance of thermal imaging for nighttime operations and tailors the application specifically to surface vessel scenarios. The review result is summarised in Table 1.

The analysis reveals that most methods rely on optical sensors, due to their ability to classify targets. Primarily, RGB cameras are used in conjunction with powerful deep learning-based detection algorithms. Some works, however, note that such systems are less effective or even completely useless in nighttime or low-visibility scenarios. UAVs are often proposed for SAR, but they have significant limitations considering

Table 1
Summary of related work on human-in-water detection systems.

Reference	Platform	Sensor	Classification
(Gennarelli et al., 2022)	N.D.	Radar	No
(Mansor et al., 2021; Jian et al., 2017)	ASV/USV	Sonar	No
(Akbar et al., 2022; Kang et al., 2020)	ASV/USV	RGB camera	Yes
(Taipalmaa et al., 2024; Wang et al., 2023; Ancy Micheal and Sivaramakrishnan, 2024; Lygouras et al., 2019; Rizk et al., 2023)	UAV	RGB camera	Yes
(Li et al., 2021)	N.D.	RGB camera	Yes
(Cheong et al., 2024)	N.D.	IR camera	No
(Feraru et al., 2020)	UAV	IR camera	Yes
(Martins et al., 2013)	ASV	IR/RGB camera	No
(Katsamenis et al., 2020)	Ship	IR/RGB camera	Yes

the scenarios detailed in Section 1. When detection is carried out on thermal images, many negative effects are mitigated, but these works only focus on detection and fail to localize the target effectively. In other cases, more effective 3D sensors such as sonar and radar are used, which are less affected by environmental disturbances. However, these sensors only provide detection, without the ability to classify the target. LiDAR sensors, although widely used for USV/ASV navigation, do not seem to be chosen for human detection in SAR operations, as they also fail to effectively classify the target, which is crucial in disaster scenarios crowded with possible debris.

In other words, there is a prevailing tendency in the state of the art to entrust detection to isolated sensors or platforms whose performance degrades significantly under the realistic SAR conditions outlined in the research reported in Section 2. Therefore, this paper aims to develop and preliminarily validate, using a relevant dataset, a multi-modal sensor fusion pipeline that combines LiDAR and thermal imaging. The proposed approach is explicitly designed to enable robust human detection in water during naval disaster scenarios, with a particular focus on resilience to environmental disturbances.

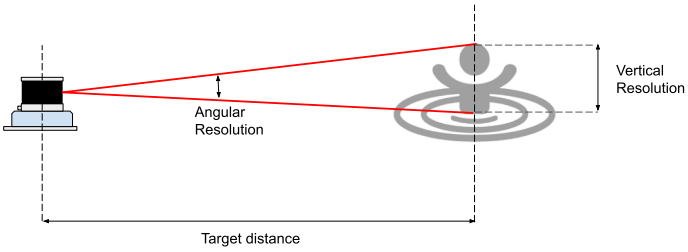
4. Material and methods

This section provides details on the methodology adopted and the material involved in the study. In particular, Section 4.1 provides an overview of the methodology and an in-depth analysis of the main steps of the study. Section 4.2 provides details on the sensors and acquisition setup, while Section 4.3 presents the hardware and software specification of the computing module. Finally, Section 4.4 shows the collection of an ad-hoc dataset.

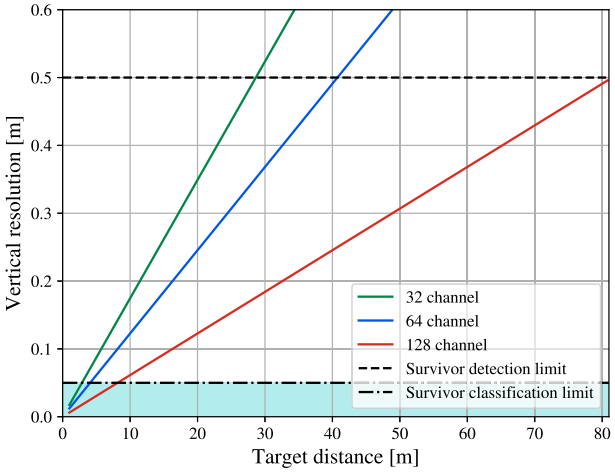
4.1. Methodology

To select an appropriate sensor setup and processing pipeline, we must first consider the application's requirements and constraints. As SAR-oriented ASVs are likely to be already equipped with LiDAR for collision avoidance and path planning (Faggioni et al., 2022a,b; Ponzini et al., 2024, 2025; Thombre et al., 2022; Clunie et al., 2021; Helgesen et al., 2022; Stanislas and Dunbabin, 2019), we will first consider the suitability of LiDAR for human-in-the-water detection. In general, the primary limitation of LiDAR-based obstacle detection lies in its vertical resolution, which is typically much lower than the horizontal resolution. A survivor's body will reliably produce at least one reflection only if its vertical dimension is equal to or exceeds the sensor's vertical resolution (see Fig. 1(a)). For a floating survivor raising their arms (with a vertical span of approximately 0.4–0.6 m), the maximum detection range for a typical LiDAR system with a 40° vertical field of view and 32–128 vertical channels is roughly between 30 m (in the best case) and 80 m (in the worst). As an illustrative example, Fig. 1(b) shows the relationship between vertical resolution, target distance, and number of sensor channels; a vertical FOV of 40° is used for the analysis, which is a typical value for many commercially available LiDARs.

While this calculation indicates the resolution required to detect any signal at a certain distance, it is unlikely that a lidar-based detector will successfully distinguish a person from the surrounding sea or flotsam based on a single point. To obtain semantic information on the target, a larger number of points must be acquired to ensure adequate feature extraction (e.g., distinguishable head shape), thus it is estimate the required vertical resolution necessary to detect a head spanning 22-24cm vertical to be at least 0.1m (considering a head vertical span of



(a) Vertical resolution limit scheme.



(b) Vertical resolution on target distance.

Fig. 1. LiDAR acquiring a survivor.

22–24cm). Using the relationship illustrated in Fig. 1(b), classifiable targets are obtained at 5 to 10m for common high-spec commercial lidars with 32–128 vertical channels. Consequently, a LiDAR-only equipped ASV will need to check individual detection targets at very close range to be able to classify them as human-in-water or floating debris. This limits the usefulness of a lidar-only ASV because of its long path along all candidate targets and correspondingly long time-to-rescue.

The addition of a camera has the potential to significantly increase the distance at which targets can be detected and classified due to the much higher resolution it offers. Additionally, a wealth of academic research and pre-trained detection models is available for the object detection problem, serving as a starting point for camera-based human-in-water detection. RGB cameras, however, are ineffective at night and under low-visibility conditions, which are commonly encountered in SAR missions, as detailed in Section 2. As reported in Section 3, several works propose the use of thermal imaging because of its superior robustness to adverse weather conditions and nighttime performance.

While thermal imagers in general have lower resolutions than their RGB counterparts, compared to LiDAR, they still provide a 5 to 20 times higher sampling density suitable for trained detectors even at a longer distance. This enables one-shot detection and classification of survivors at greater distances, even in cluttered environments, thereby reducing the average path length to reach the rescue point and, consequently, the time-to-rescue.

Since thermal imagers do not directly estimate distance, any detection must be coupled with detection via a 3D sensor (such as LiDAR or radar) to obtain the range to the target and formulate a safe route to reach it. Considering the difficulty of maritime radar detection close to the sea surface and the reliance of mainstream ASVs on LiDAR for obstacle avoidance, we propose a combined system that uses thermal imaging and LiDAR for long-range person-in-water detection and localization.
A simple and widely used method for combining the detection of two sensors, such as a thermal camera and LiDAR, is late fusion, where detections are made by the individual sensors and then associated at the end of the pipeline. While such a system combines the desirable attributes of thermal detection and LiDAR ranging, its detection performance can never exceed that of the thermal camera, as its processing pipeline does not use any additional information obtained by the LiDAR to strengthen its accuracy this aspect will be better clarified in Section 5.1.

Additionally, thermal imaging still suffers from some degrading environmental conditions that are generally rare, but become more likely in a maritime disaster scenario, as outlined in Section 1. In Rivera Velázquez et al. (2022), the authors investigate the decay in detection and classification performance on thermal images with the presence of fog and haze. The Normalized Global Contrast (NGC) of a target with a temperature close to the human body can be decreased by 20 to 60 % by a 25m Meteorological Optical Range (MOR) and up to 65–75 % for a 15 m MOR. The MOR decreasing can be caused by haze/fog or rain due to adverse weather, but also by smoke caused by the combustion of fuel spill or wreckage, and by engine failures and other machinery malfunctions; in addition to smoke, the presence of fires in the area provides the possibility of anomalous hot zones. Additional anomalous bright spots in the image can be produced by atmospheric phenomena, especially near the horizon as observed in the MassMIND dataset (Nirgudkar et al., 2022) and reported by Cheong et al. (2024). In adverse weather conditions, the thermal camera is susceptible to spray occlusion due to green seas, which, depending on the temperature, can produce cold or hot localized noise spots. Finally, sensor temperature anomalies and humidity infiltration can compromise the image quality.

We hypothesize that, despite the limited resolution of the LiDAR, its sparse point cloud still provides useful geometric clues that can aid the thermal processing pipeline in identifying object outlines and sizes. Multi-modal early-fusion, in which features are extracted from joint thermal and LiDAR measurements, could provide better final system performance due to the partially non-overlapping susceptibilities to environmental conditions. LiDAR information can compensate for thermal failures in the presence of haze and noise. Additionally, LiDAR can help avoid false positives related to horizon bright spots due to the upper limit on signal reflection distance. Thermal imaging, on the other hand, can compensate for LiDAR's low point density, enabling classification even at high distances with a limited number of points acquired on the target.

In summary, early-fusion multi-modal detection can be more resilient to challenging environmental conditions and soft sensor failure scenarios.

Considering all these factors, a human-in-water detection pipeline is proposed based on early fusion of thermal imaging and LiDAR data. A general scheme of the SAR scenario is provided in Fig. 2, where the sensing layer, composed of LiDAR and a thermal camera, acquires the point cloud and thermal image of a survivor at sea surrounded by floating debris.

In an early fusion method, the information of two modalities is combined at the data level, prior to feature extraction. Care must be taken to preserve the spatial relationships between heterogeneous data, as robust network architectures, such as convolutional neural networks (CNNs), rely on local feature extraction followed by aggregation in a scale pyramid approach. Therefore, the LiDAR point cloud is first crop based on the camera's Field of View (FOV) and, through LiDAR-camera calibration, it is projected onto the image plane. From the resulting 2D-projected point cloud, two LiDAR-derived images are generated using the reflectivity and range fields. The two LiDAR-derived images are stacked with the thermal image to form a 3-channel multi-source image, within which the human-in-the-water is detected using a YOLO-based neural network. The 3D coordinates of the detected survivor(s) are then extracted, and a target point is identified based on relative distance and heading angle, enabling precise localization, a crucial information for computing a possible rescue route. Fig. 3 illustrates the overview of this detection and classification pipeline. The following subsections provide a detailed explanation of the individual processing steps.

4.1.1. Calibration

To enable LiDAR and IR camera sensor fusion, it is necessary to obtain the intrinsic and extrinsic calibration parameters, as well as a distortion model. Therefore, a calibration procedure using the Zhang (2000) method was performed. Various methods for calibrating thermal cameras are available in the literature, utilising different patterns and techniques to make them visible through thermal imaging. For this work, a stationary checkerboard was heated by a halogen lamp. A 5-parameter polynomial model was used for the distortion parameters, considering both radial and tangential distortion. The extrinsic parameters were obtained from the technical drawing of the sensor mounts, followed by manual refinement of the rotation angles to achieve good alignment between thermal and point cloud object boundaries. The perspective projection of LiDAR points onto the image plane follows the pinhole camera model as shown in (1).

$$s \cdot \mathbf{p} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]\,\mathbf{P} \tag{1}$$

where:

- $\mathbf{P} \in \mathbb{R}^{4\times1}$ is the 3D point in homogeneous coordinates, i.e., $\mathbf{P} = [X, Y, Z, 1]^\top$;
- $\mathbf{p} \in \mathbb{R}^{3\times1}$ is the projected image point in homogeneous coordinates, $\mathbf{p} = [u, v, 1]^\top$;
- $\mathbf{K} \in \mathbb{R}^{3\times3}$ is the intrinsic camera matrix derived from the calibration procedure;
- $\mathbf{R} \in \mathbb{R}^{3\times3}$, $\mathbf{t} \in \mathbb{R}^{3\times1}$ define the extrinsic transformation from the LiDAR frame to the camera frame;
- $s \in \mathbb{R}$ is a projective scaling factor.

4.1.2. Thermal-LiDAR 3-channel image generation

To create the multi-source 3-channel image, the LiDAR point cloud must be projected onto the image plane of the thermal sensor. As a first

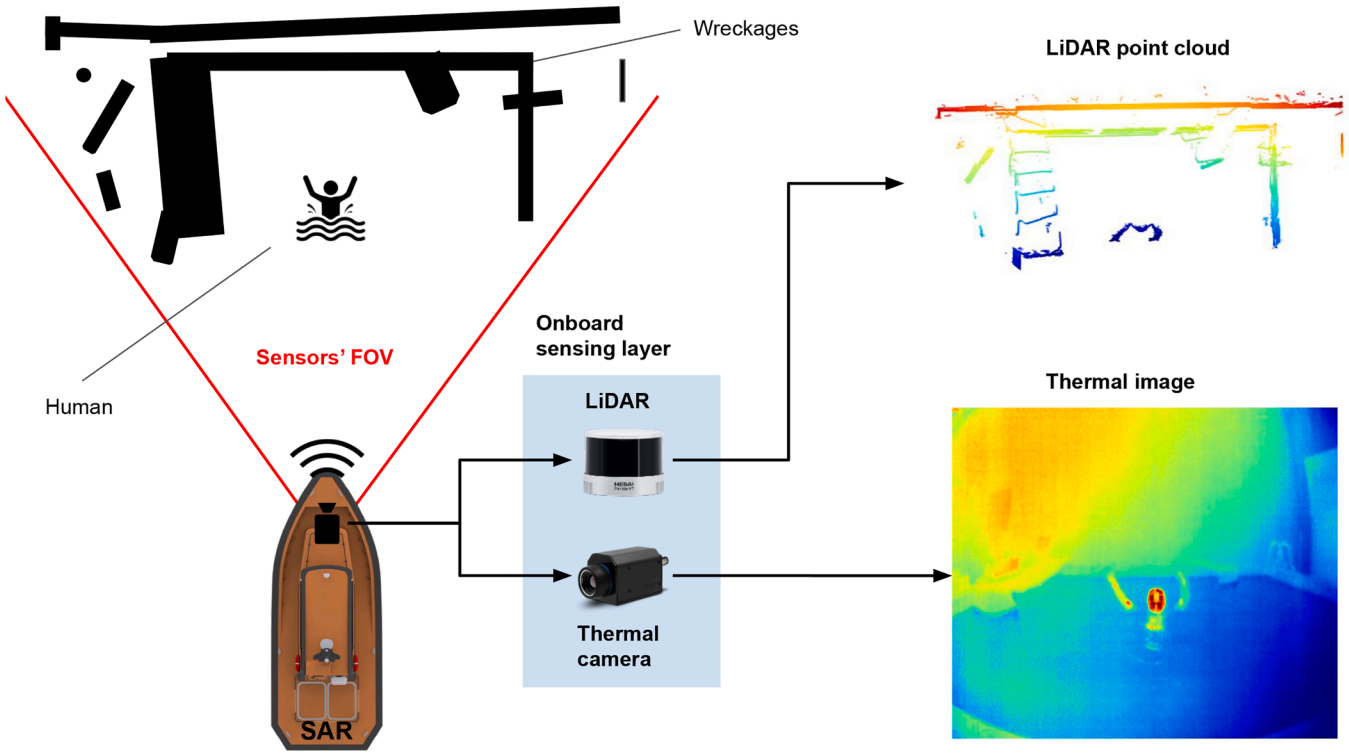


Fig. 2. Operative Search And Rescue Scenario with sensing layer outcome.

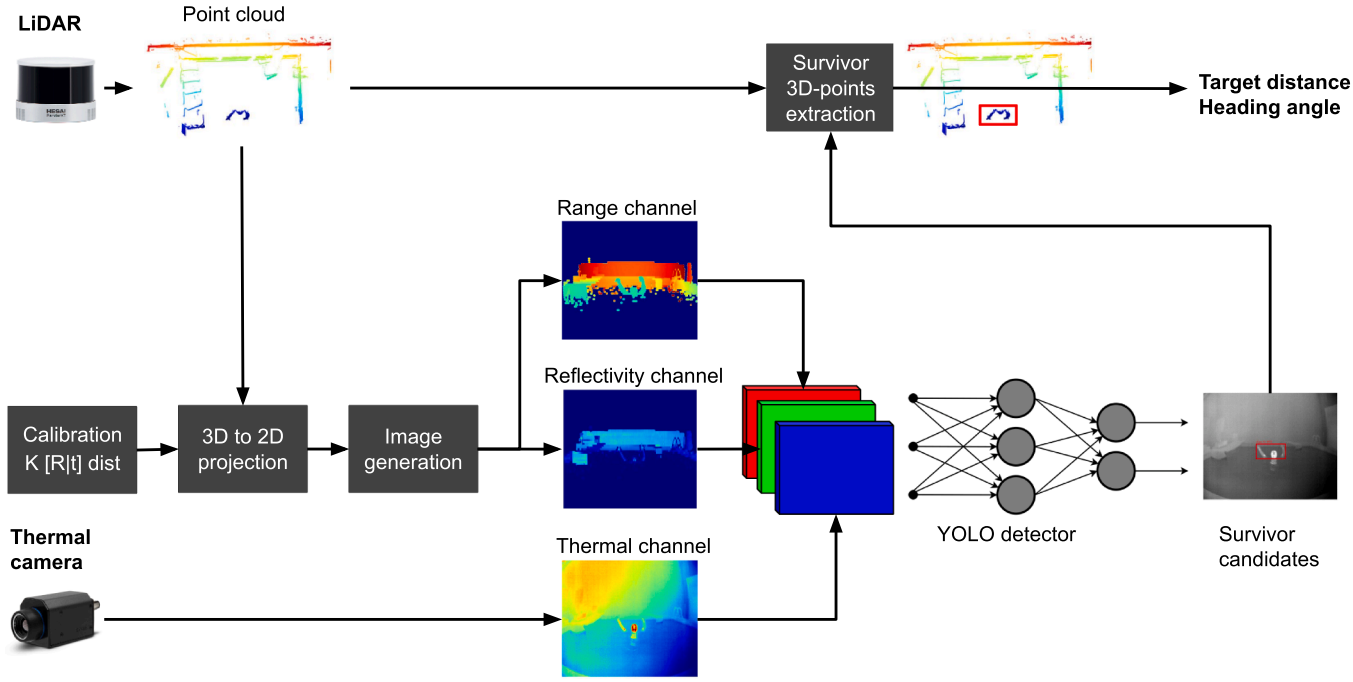


Fig. 3. Early fusion LiDAR thermal camera pipeline.

step, the LiDAR point cloud is cut to the FoV of the thermal camera. Secondly, the 3D point cloud is transformed to the 3D camera coordinate system of the camera using the extrinsic calibration angles and translation. Thirdly, the thus transformed point cloud is subjected to the pinhole camera perspective projection using the intrinsic parameters obtained from the thermal camera calibration. This provides a sparse point-to-pixel mapping mechanism that relates the 3D lidar information to 2D positions in the (undistorted) thermal camera image. Two LiDAR images of the same resolution as the thermal image are generated by filling pixels with the reflectivity and range values of the corresponding LiDAR point, respectively, normalized to the 0–255 interval.

Nearest-neighbor interpolation (within a maximum distance determined by the resolution disparity between sensors, up to 8 pixels) is used to convert the sparse LiDAR projection images to dense images, filling in the gaps between the projected points.

(a) Original point-cloud projected on the image plane. (b) Field-map image without interpolation. (c) Interpolation action on the field-map image.

Fig. 4. LiDAR field-map generation process.

The thermal image is stacked together with the two LiDAR-derived (Reflectivity, Range) images to form a multi-source image enriched by the features extracted from the two sensors. To provide a visual representation of the procedure, Fig. 4(a) shows the LiDAR point cloud acquired from a dataset scenario, superimposed on the corresponding thermal image. Fig. 4(b) displays the same image without Nearest Neighbor interpolation, and Fig. 4(c) illustrates the result after applying the interpolation. The thermal image encodes information about object temperature, the LiDAR reflectivity channel provides cues about material properties, and the LiDAR range field captures distance. Combined, these modalities help effectively distinguish objects based on temperature, distance, reflectivity, or any combination of these factors.

4.1.3. Detection and classification

To perform one-shot detection and classification of the survivor(s) in the multi-sensor image, the YOLOv8 detection CNN (Jocher et al., 2023) is adopted. YOLO (You Only Look Once) is chosen because it offers an excellent trade-off between accuracy and execution time, and is easily scalable across different compute platforms, as multiple variants with varying memory and compute requirements are provided.

No modifications were made to the model architecture, as its standard configuration already supports 3-channel inputs. However, since the available pre-trained weights are tailored for RGB images (not suitable for the case study), the model was retrained from scratch. The YOLOv8 detector is trained and validated on an experimentally acquired 3-channel multi-source image dataset, augmented to include a sufficient number of adverse conditions (details in Section 4.4).

4.1.4. Rescue points extraction

The YOLO CNN provides image plane bounding boxes for human-in-the-water detections, delineating their extents within the image. Using the point-to-pixel mapping described above, the image bounding box can be related to a cluster of corresponding 3D LiDAR points that project within this bounding box, without requiring a specific clustering analysis on the 3D point cloud and therefore saving computational time.

In the case of objects superimposed on the human figure, the bounding box may also contain 3D points from the environment in front of or behind the detected person. This special case can be automatically detected by comparing the range variance with the typical span of a human figure, and if detected, adaptive thresholding on the distance histogram inside the bounding box is performed to separate the human cluster from the background cluster. To identify the cluster belonging to the person in the water, the median value of the thermal intensity can be computed. Alternatively, the mean reflectivity of a 3D cluster can be compared to the median reflectivity value of the LiDAR image.

Table 2
LiDAR main specifications.

Spec.	value
Vertical FOV	31°
Vertical resolution	1°
Horizontal FOV	360°
Horizontal resolution	0.09° to 0.36°
Operating Frequency	5 Hz to 20 Hz
Range	120 m
Accuracy	0.01 m

The rescue point extraction provides the relative heading angle and distance concerning the sensor system. These output variables are critical for initiating reactive maneuvers and supporting path planning modules, as outlined by Zaccone (2024).

4.2. Sensors

To carry out the experimental data acquisition campaign and to develop and test the pipeline with real-world data, a dedicated acquisition setup was constructed comprised of a LiDAR sensor, a thermal camera, and a custom-designed aluminum alloy support, which can be mounted either on a tripod via a dedicated insert or on a vehicle using a bolted bracket. The LiDAR used is the 32-channel HESAI Pandar XT-32, as it represents an affordable mid-range LiDAR, and its performance is already extensively validated in the field, both in port areas and in blue water scenarios, as demonstrated in previous works (Faggioni et al., 2022a,b; Martelli et al., 2022). Its technical specifications are detailed in Table 2, while Fig. 5 illustrates the principal dimensions and the sensor's channel distribution.

The thermal camera selected for the setup is the Teledyne FLIR A65 (see Fig. 6), which offers a vast Field of View (FOV), making it well-suited for data fusion with a LiDAR sensor. This camera was selected as it ensures a good trade-off between accuracy and field of view, further supporting its integration into the sensing layer; the thermal camera technical specifications are provided in Table 3. Finally, an overview of the support structure is shown in Fig. 7. In Fig. 7(a), the support structure is presented along with the respective sensor installation dimensions, while Fig. 7(b) displays a photo of the acquisition system.

4.3. Computing hardware and software

The acquisition, processing, and testing activities were performed on an off-the-shelf computer with the hardware specifications reported in Table 4. The pipeline was implemented in Python 3.11 with GPU support through PyTorch 2.4.1, plus ROCm 6.0.

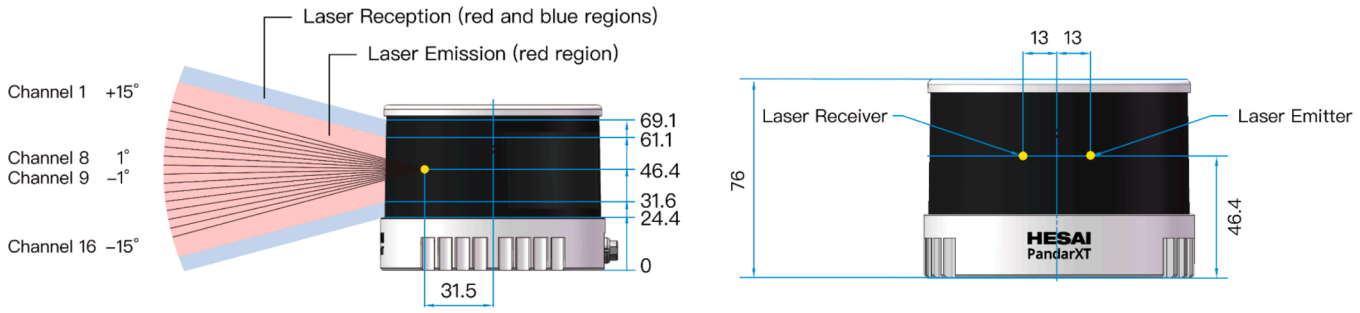


Fig. 5. HESAI Pandar XT-32 LiDAR technical drawings.

Fig. 6. Teledyne FLIR A65 thermal camera.

Table 3
Thermal camera main specifications.

Spec.	value
FOV	90°×69°
f-number	1.25
Image Frequency	30Hz
Detector Pitch	17 μm
Focal Length	7.5 mm
IR Resolution	640×512 pixels
Spectral Range	7.5–13 μm
Thermal Sensitivity	< 0.05°C @ 30°C
NETD	50mK
Accuracy	±5°C

Table 4
Personal computer main specifications.

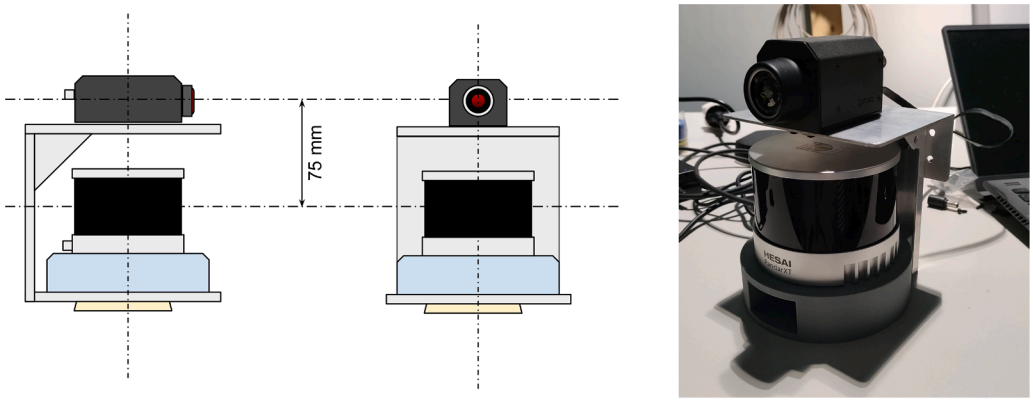
Spec.	Value
CPU	AMD Ryzen 9 6900HS 4.9 GHz
RAM	16 GB DDR5-SDRAM 4800 MHz
GPU	AMD Radeon RX 6700S
VRAM	8 GB GDDR6

Table 5
Dataset physical augmentation.

Category	Item
Floating objects	Generic debris Flotation aid Life-jacket
Human actions	Cling to tank-side Cling to debris Cling to flotation aid Raise arms Submerge head Float passively Float actively
Human wear	Wet-suit Life-jacket

4.4. Dataset

A typical challenge of computer vision approaches in the maritime field is the scarcity of open-access data, which becomes even more pronounced in applications such as the one proposed. To address this, several experimental data acquisition sessions were organized using the setup described in Section 4.2. These sessions were conducted in a controlled indoor environment at the COMPASS Lab (University of Genoa, La Spezia campus). The laboratory is equipped with a test tank for testing Guidance, Navigation, and Control (GNC) architectures of ASV/USV (details can be found in Ponzini et al. (2023)), which in this case was used to recreate the human-in-water scenarios. In particular, a person was made to float in the tank and imaged by the acquisition system from various positions. Additionally, to represent a variety of SAR scenarios, different variations of the scenario were created: the person either wore



(a) Sensors set-up, technical drawing. (b) Sensor set-up, real image.

Fig. 7. LiDAR - thermal camera set-up.

or did not wear a life jacket, floating debris and flotation aids were scattered in the water, the tank's edges were masked with surfaces simulating a wreck, and the person enacted different levels of activity. Among the actions performed were submerging the head to reduce heat, floating in a passive position (as a "dead man"), clinging to debris or flotation aids (such as an inflatable buoy), gripping the edge (or wreck) of the tank, and raising arms to simulate a distress signal during drowning. Table 5 summarizes all the actions carried out to complicate the scenario, categorized accordingly; it should be noted that these actions were randomly mixed to create a distributed and generalized dataset. The dataset thus acquired provides 3531 pairs of thermal images and point clouds.

During the acquisition, both images and LiDAR point clouds were saved along with their respective timestamps, using custom-developed acquisition codes. To achieve the maximum horizontal resolution, the LiDAR was set to acquire data at a frequency of 5 Hz. Consequently, it was necessary to synchronize the images with the point clouds using the timestamps, resulting in pairs of point clouds and thermal images extracted at a frequency of 5Hz. Synchronization was achieved by associating each LiDAR point cloud with the image bearing the closest timestamp. The images were manually annotated using a custom-developed tool based on the OpenCV library, following the YOLO format. Bounding boxes were drawn around all visible human parts emerging from the water, ensuring accurate localization of the human subject for training purposes. Each synchronized image–point cloud pair was saved in a single JSON file, accompanied by a TXT file containing the annotation in YOLO format. The corresponding Range and Reflectivity LiDAR images are generated on-the-fly, during testing, using a dedicated procedure based on perspective projection and Nearest Neighbor interpolation. This approach is essential to faithfully replicate real-time operational conditions during post-processing, particularly in the analysis of computational costs presented in Section 5.3.

To increase the number of samples, including potential adverse weather conditions, as outlined in Section 2, the dataset was augmented to account for various weather phenomena, such as rain, green seas, fog/smoke, noisy bright spots, etc. The weather phenomena along with their respective augmentations are listed in Table 6. While the augmentation process is inherently synthetic, care was taken to preserve physical plausibility. Specifically, fog and smoke effects were modeled based on the reduction of Normalized Global Contrast, as described by Rivera Velázquez et al. (2022). To simulate spray and green sea conditions, the pixel intensity of the water surface was extracted from reference images and reintroduced into the augmented data, with a random variation of up to ±50 % to account for possible temperature fluctuations. Note that the water temperature during testing was 19°C. The addition of bright spots was modeled by analyzing their potential sources and appearance: (i) based on observations across multiple images such as in Cheong et al. (2024), Nirgudkar et al. (2022); (ii) by taking into account reflections on the water surface, as directly observed in our data and illustrated in Fig. 8; (iii) by considering the camera's operating temperature (observed in 5 – 10°C above room temperature), which can produce localized heating on droplets deposited on the lens, thus generating spot-like artifacts. In addition to these, standard augmentations such as flipping and distance scaling were performed. To produce realistic distance-scaled LiDAR point clouds, point density was reduced in addition to scaling the point range values. Typical noisy LiDAR points, caused by water surface ripples, were naturally present due to the motion of the survivor actors in the tank.

Well-synchronised image–thermal LiDAR point cloud pairs are then processed to obtain the projection of the point cloud onto the image plane, along with the two resulting LiDAR-derived images of the range and reflectivity fields, according to Section 4.1. Subsequently, using a self-developed tool, each multi-source 3-channel image is labeled in YOLO format.

Fig. 9 shows a sample of the dataset subjected to the primary preprocessing step. In particular, Fig. 9(a) shows the thermal image with a rainbow colormap to represent temperature variations. Fig. 9(b) provides the Bird's Eye View (BEV) point cloud, which displays a survivor with raised arms, the test tank border, and artificial debris. Fig. 9(c) and (d) show the point cloud overlapped onto the thermal image plane, colored by reflectivity and range, respectively; these two figures illustrate the alignment of the modalities and the basis for creation of the range and reflectivity images.

Fig. 10 shows three random examples from the labeled dataset of non-augmented 3-channel images. Rows correspond to different examples, while columns represent the individual channels (thermal, reflectivity, and range). The images are displayed in grayscale format, and the bounding box enclosing the human-in-water is shown in red. The greater visual impact of the range field is due to the fact that in this channel, the differences between objects are much more pronounced than on the intensity channel. Examples of augmented thermal images are shown in Figs. 11 and 12. Fig. 11 shows progressive fog augmentation (and thus a meteorological optical range decay), which, according to Rivera Velázquez et al. (2022), is reflected in a decrease of the target-to-background contrast. Fig. 12 shows a random example of the combined effect of the other possible augmentations. From left to right, distance augmentation plus light haze, tin smoke with some water drips, sea-blast, heavy rain with wet lens. This procedure enables the creation of a dataset augmented as desired to include environmental disturbance phenomena of varying aggressiveness, thereby increasing the number of samples useful for training the YOLO network. For the case study, the original un-augmented dataset was divided into 80 % training and 20 % validation. From the training dataset, 6000 randomly augmented multi-source images were generated, creating a dataset divided into samples with no augmentation, mild augmentation, and substantial augmentation, to account for the broadest possible range of conditions. Three groups with increasing augmentation aggressiveness were generated from the validation dataset. It should be noted that, due to the prior separation of the source data, the model is never exposed to the validation dataset during training, nor is the validation source data used for generating augmented training images.

Table 6
Thermal image meteorological augmentation.

Phenomenon	Effect	Augmentation
Fog	Hazed image	Weighted mask and/or blur
Thick smoke	Heavily hazed image	Strong gradient
Green seas/spray	Water drops	Small-sized random cold spot
Horizon lights	False positive bright spots	Medium-sized random hot spots

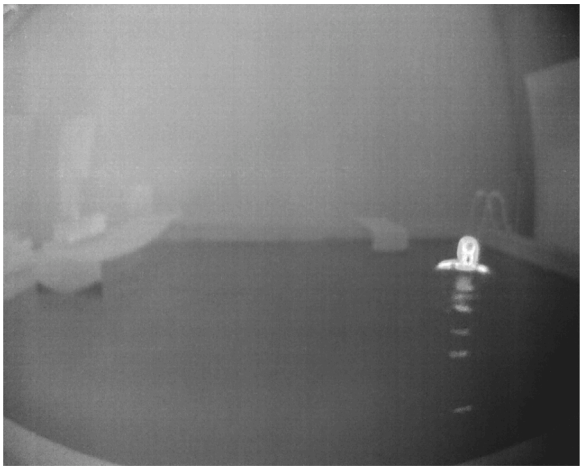
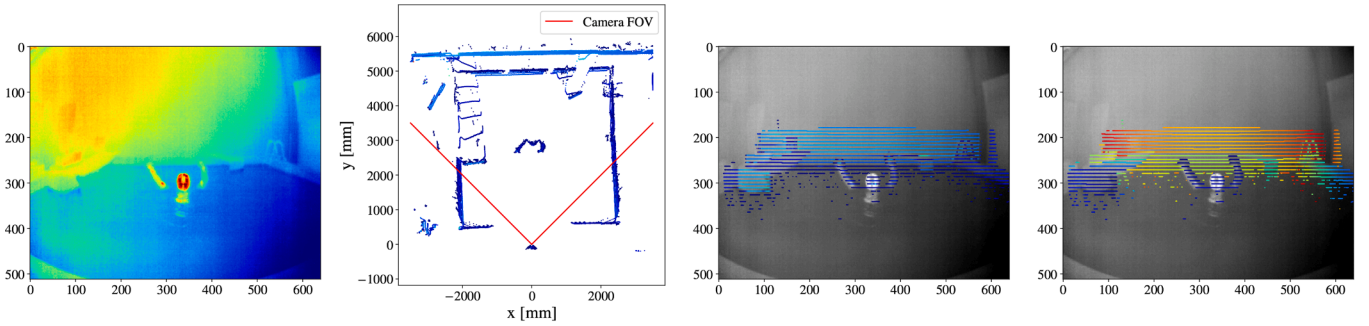


Fig. 8. Bright spot reflection on water surface observed during the tests.



(a) Reference thermal image. (b) BEV LiDAR point cloud (c) Projected point cloud coloured by reflectivity. (d) Projected point cloud coloured by range.

Fig. 9. Preprocessing step on the LiDAR point cloud.

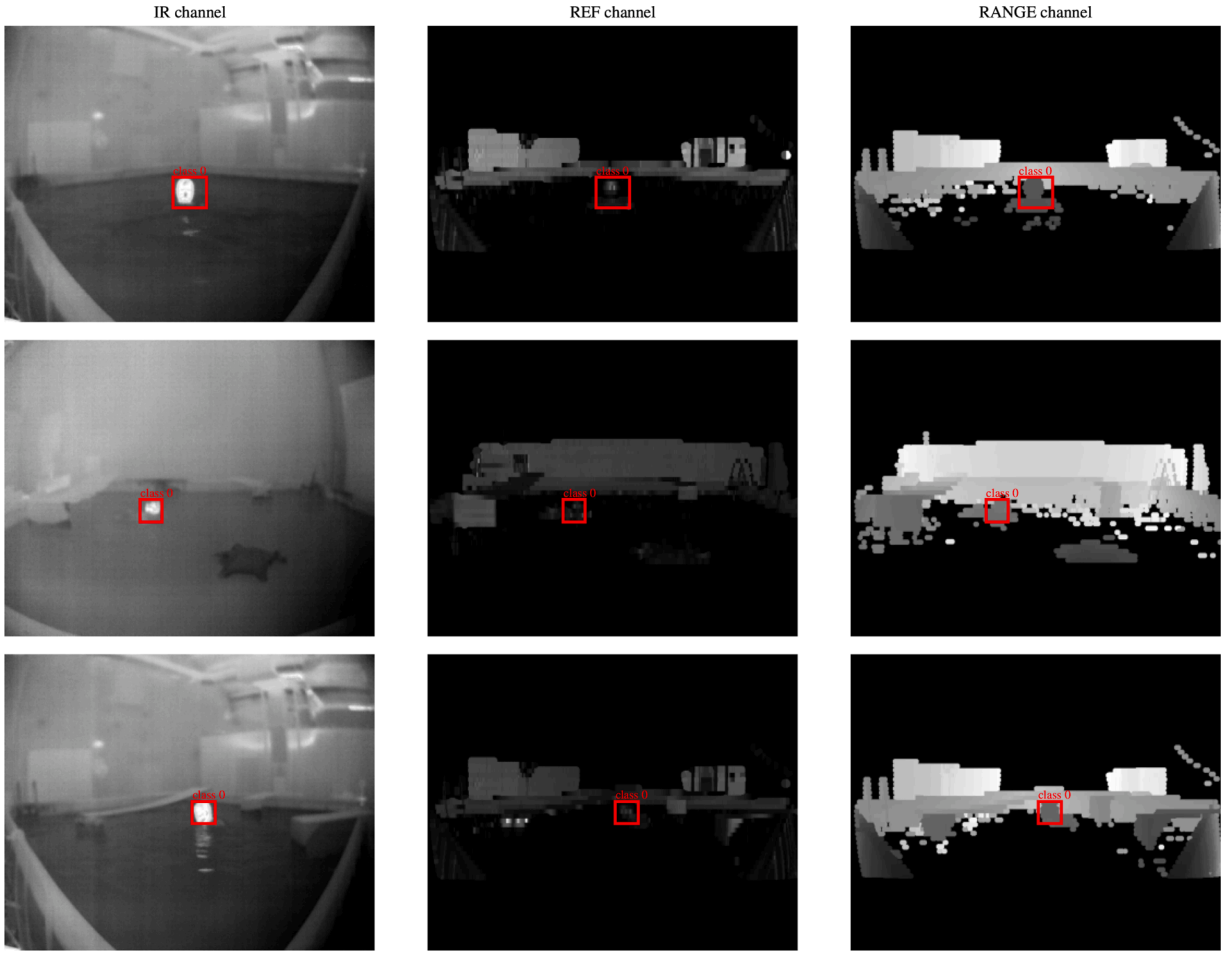


Fig. 10. Shuffle samples of the labeled 3-channel images.

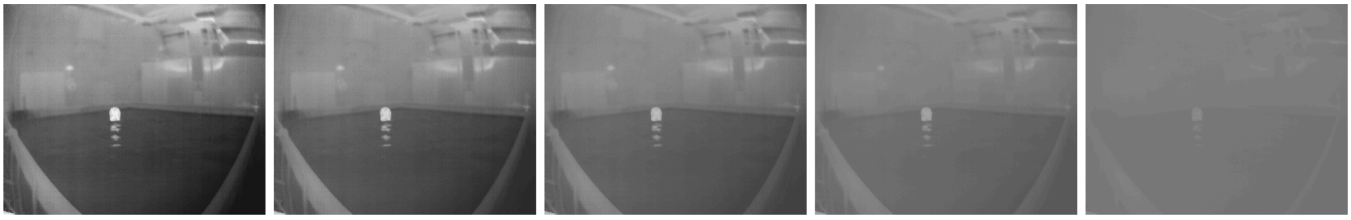


Fig. 11. Fog augmentation with progressive Meteorological Optical Range decay.

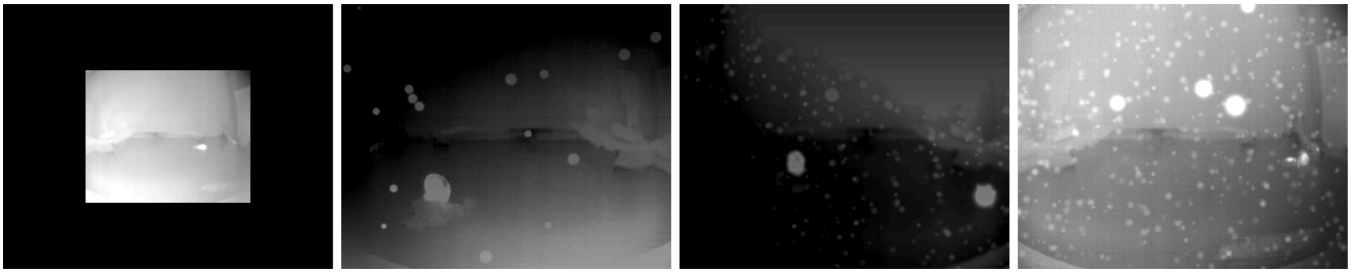


Fig. 12. Random augmentations.

5. Results

This section presents the results obtained. In particular, Section 5.1 presents a preliminary analysis that supports the use of sensor fusion and the training and validation metrics of the YOLO network. Section 5.2 displays the results of the complete pipeline, from multi-sensor acquisition to the generation of the target point for the rescue path. The obtained results are also compared against single-modality detection approaches and/or late-fusion strategies. The late-fusion approach can be implemented in two main ways: (i) image-based detection with LiDAR spatial characterization, where detection and classification are performed solely on the image data (e.g., thermal), and spatial information from the LiDAR is incorporated via perspective projection. In this case, the detection performance is equivalent to that of the thermal image processing branch (or the single modality), with the only difference being the spatial characterization (see Section 5.1); (ii) independent detection and association, in which a detector is applied to the thermal data while a separate processing pipeline (typically clustering) is applied to the LiDAR data. The resulting detections are then associated. However, the performance of this approach is highly sensitive to parameter selection, association strategy, and candidate confirmation criteria, which makes an objective performance comparison challenging. For this reason, in this work, we opted to monitor only the computational cost of this pipeline (see Section 5.3).

5.1. Rough weather influence

In addition to the widely discussed motivations presented in Sections 2 and 3, the authors propose a further analysis to support the necessity and benefits of a multi-modal approach for human-in-water detection. Specifically, a YOLOv8 network is trained using only thermal images, while another is trained on the 3-channel multi-source images. The base training dataset consists of the original data with standard augmentations only (vertical flip, horizontal flip, distance), excluding any augmentations related to environmental disturbances, for a total number of 6000 samples, referred to in this work as *Original*. Two levels of severity of adverse condition augmentation are then applied on this original dataset, referred to as *Augmented* and *Augmented+*, respectively. The thermal-only and thermal-LiDAR YOLOv8 detectors are trained exclusively on 80 % of the Original dataset and tested on the unseen 20 % of the Original, Augmented, and Augmented+ datasets. This is repeated for 5 non-overlapping 80-20 splits of the dataset in a standard 5-fold cross-validation scheme. Fig. 13 presents the average results across the folds in terms of precision, recall, mAP50, and F1-score. The two models (thermal-only and thermal-LiDAR) perform similarly when tested on the *Original* dataset; however, as the level of adverse condition augmentation increases, the performance of the multi-source model degrades less than that of the thermal-only model. This demonstrates that although both models are trained on the standard dataset, the multi-modal detector generalizes better to challenging conditions not present in the training. This suggests that the multi-modal detection model has learned to extract redundant and complementary features from the two modalities, increasing resilience and maintaining useful performance even on harshly augmented data samples. Following the same evaluation scheme, the results are presented in terms of the average number of false positives and false negatives per image on the validation set, as a function of the augmentation level. These results, illustrated in the Fig. 14, are particularly relevant in Search and Rescue scenarios, where the consequences of false detections are critical, either by incorrectly identifying a location as a rescue target, or by failing to detect an actual human in need. Based on the analysis of average false positives per image, it is clear that the multi-modal model, benefiting from LiDAR input, is better able to distinguish between a valid target and similar false ones. The reversal in the trend of false positives is likely due to the fact that, in the *Augmented+* set, the disturbances become so severe that, although detection performance further deteriorates (as shown in Fig. 13), the features available for detection are significantly reduced; this scarcity of features also impairs the model's ability to generate false positives, even though the average number of false positives remains high and relatively stable. Finally, this result highlights how the early fusion approach (which uses the multi-modal detector) allows greater resilience to adverse weather conditions compared to an image-based late-fusion approach (which would use the thermal-only detector).

To evaluate the detection and classification performance of the YOLO network on multi-source images, it is performed a k-fold cross-validation using a general-purpose dataset (*Augmented*) that contains a balanced mix of non-augmented, mildly augmented, and aggressively augmented images to enhance generalization. Fig. 15 presents the results in terms of precision, recall, mAP50, and F1-score. Each column refers to a k-fold permutation test. To provide a visual representation of the YOLO detector's output, Fig. 16 shows the bounding boxes obtained on the three channels for three samples from the dataset. Rows correspond to the three samples, while columns represent the channels of the multi-source image, namely thermal, reflectivity, and range.

5.2. Sensor-to-rescue results

The trained YOLO model is embedded throughout the entire pipeline, which begins with the sensor and culminates in the estimation of the human's position in water. The input data is shown in Fig. 17 within the field of view (FOV) of a hypothetical ASV/USV. Fig. 17(a) displays the point cloud in BEV, clipped to the camera's FOV, using a colormap proportional to the range; Fig. 17(b) shows the corresponding thermal image. Fig. 17(c) and (d) show the LiDAR-derived images of the intensity and range fields, respectively. Fig. 18 shows the pipeline output. Specifically, Fig. 18(a) displays the detection from the YOLO network; for visual simplicity, the bounding box is shown on the thermal image, as it is easier to interpret. Fig. 18(b) shows the ASV's navigation plane with the LiDAR points acquired in BEV (still useful for collision-free path planning) in black and the points extracted from the region of space in the image plane occupied by the human-in-water detected by the YOLO network in red. In Fig. 18(c), the target point is extracted in terms of range and relative heading angle, indicating the position of the survivor to be rescued. Note that one of the worst-case scenarios was chosen, where, as discussed in Section 4.1, the region in the image plane identifying the survivor contains, on the two LiDAR channels, points that

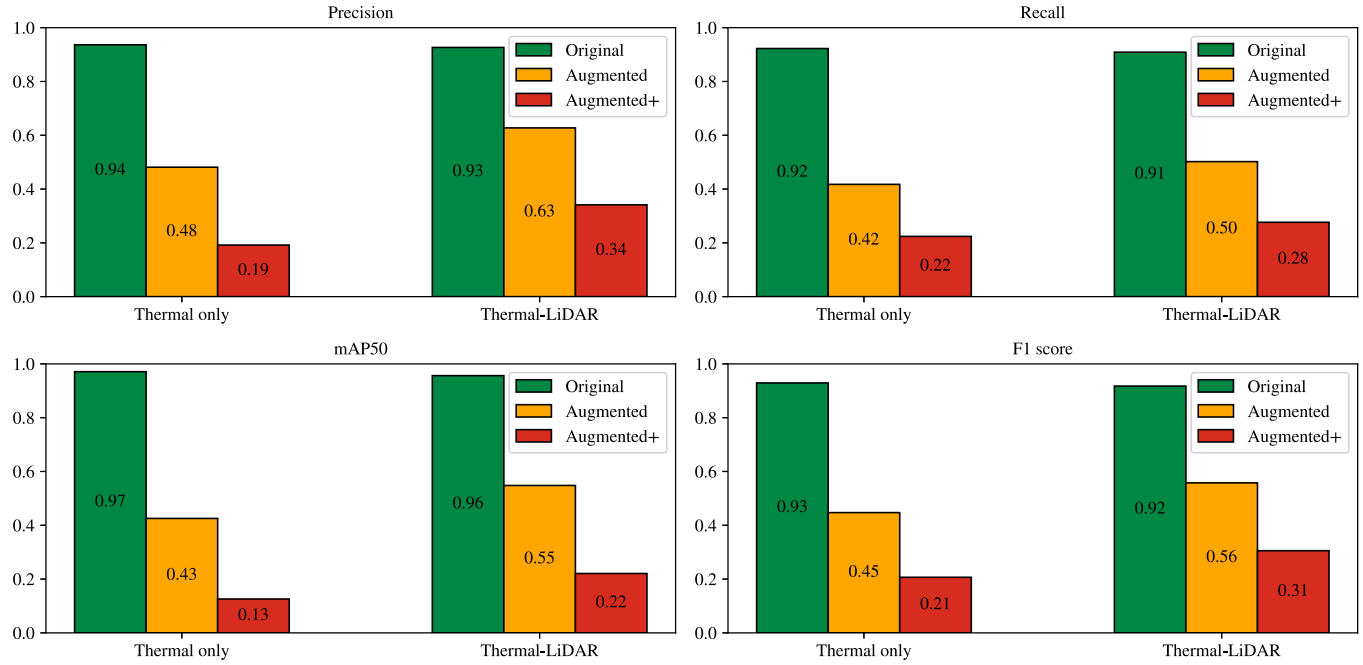


Fig. 13. Comparison of the metrics of the YOLO detector trained on thermal-only images and thermal-LiDAR combined images with varying augmentation. Note that both detectors are only trained on the *Original* samples.

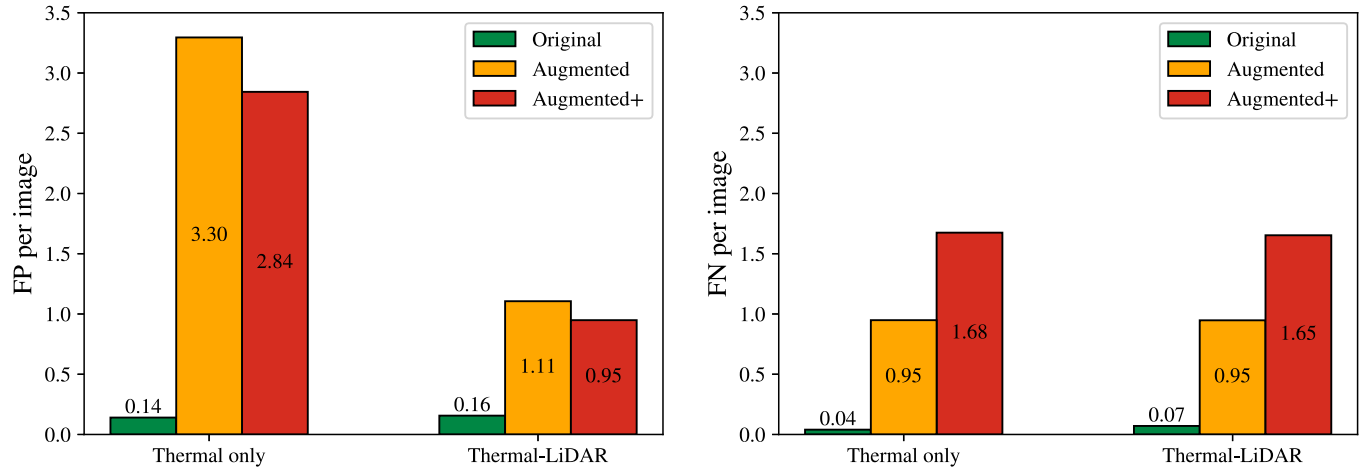


Fig. 14. Average number of false positives (FP) and false negatives (FN) per image on the validation set.

do not belong to the survivor (as can be seen in red in Fig. 18(b)); thus, the target-point estimate is challenging.

Fig. 19 shows the histogram for the three channels. In the range channel, the two space regions within the bounding box (human and debris-like object) can be identified through adaptive thresholding. The position of the pixels corresponding to these points is associated, in addition to the thresholded range value, with the thermal intensity and LiDAR reflectivity values. Using the median value of these two fields, the survivor is adequately separated from the object. Fig. 20 shows the result of this process. Fig. 20(a) shows the points belonging to the human in the water, separated through thresholding, and Fig. 20(b) shows the generation of the correct Target Point.

5.3. Computational cost analysis

To demonstrate the system's real-time compliance, a computational cost analysis was performed. Specifically, the system's processing time was monitored across 2500 scans on the system described in Table 4, with YOLOv8 detection executed on the GPU and all other processing running on the CPU.

The computational cost was divided into three main components. *Preprocessing time*: the time required to acquire data from the sensors and generate the three-channel multi-source image; *Detection Time*: the computational cost of the detection phase, performed using a YOLO-based network;*Extraction Time*: the time needed to derive the target point of the survivor in the 3D reference system from the detected bounding box. The sum of these three components constitutes the *Total time*. Fig. 21 graphically summarizes the obtained time results, while the corresponding numerical values are reported in Table 7. The time constraint considered for real-time operation is 200 ms, which corresponds to the interval between two consecutive LiDAR scans when operating at the maximum resolution frequency of 5 Hz. This threshold is depicted as a dashed black line in Fig. 21. It is worth noting that the average computational cost remains below 100 ms, thus allowing the system to operate the LiDAR at higher frequencies if necessary. However, higher frequencies may result in a decrease in spatial resolution, which could negatively impact

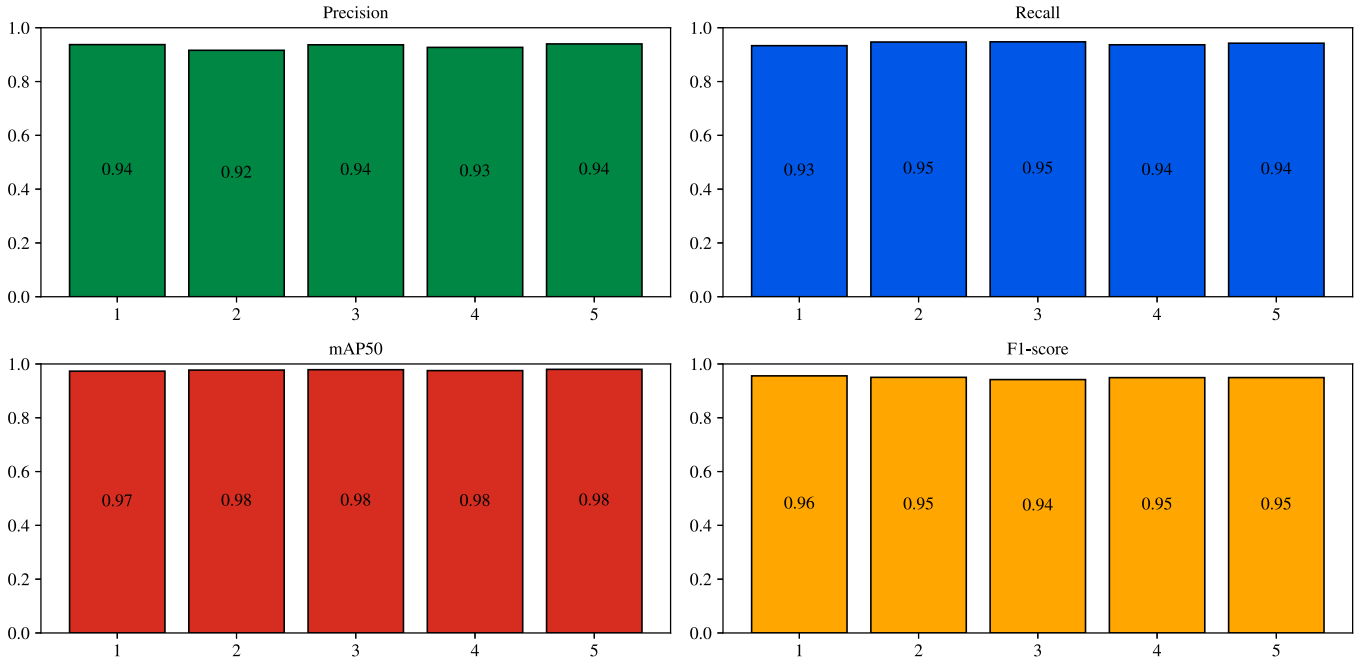


Fig. 15. 5 folds cross validation metrics results.

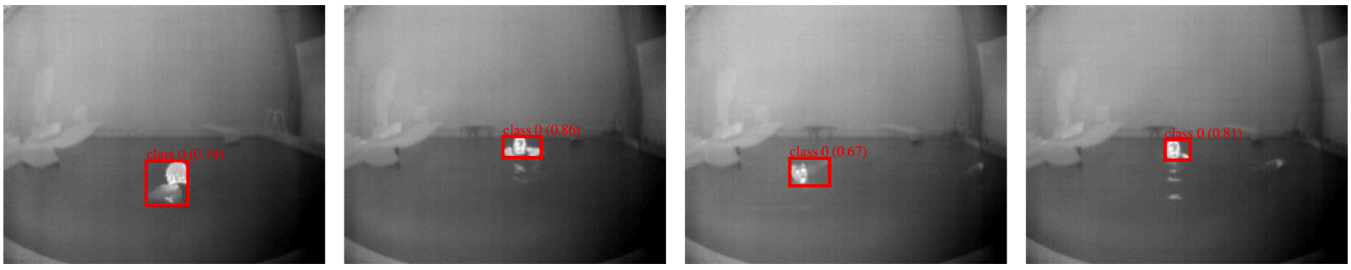
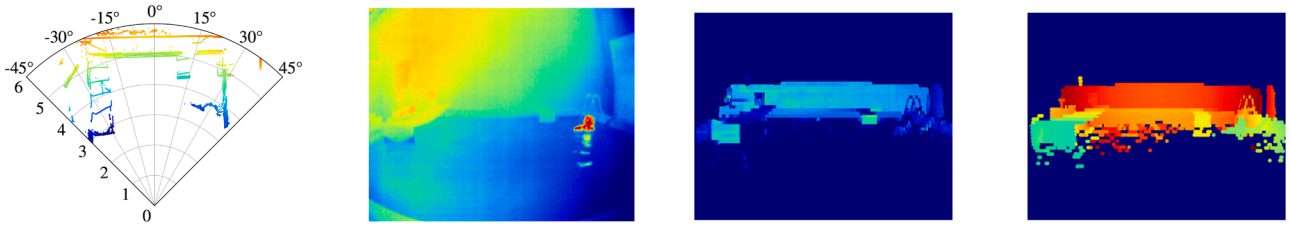
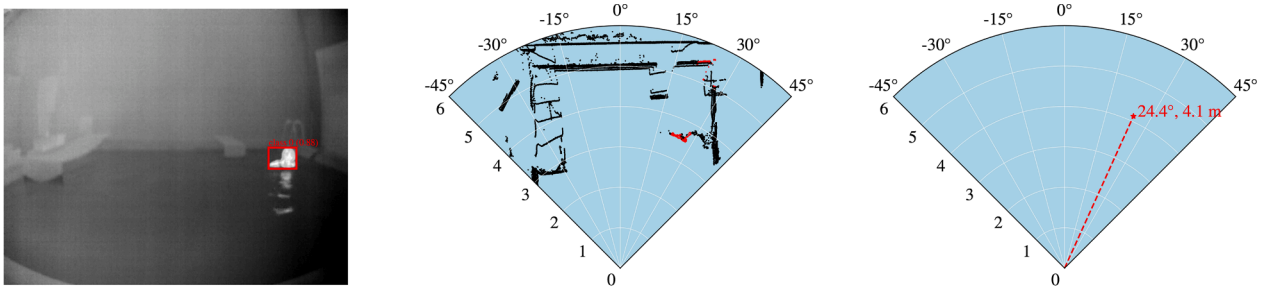


Fig. 16. Yolo detection on the multi-source image.



(a) LiDAR point cloud. (b) Thermal image (c) Reflectivity image. (d) Range image.

Fig. 17. Input data of the pipeline.



(a) Detection in the thermal image. (b) Survivor point candidates in the navigation plane. (c) Survivor candidate position in the navigation plane.

Fig. 18. Survivor detection on thermal image and LiDAR point cloud.

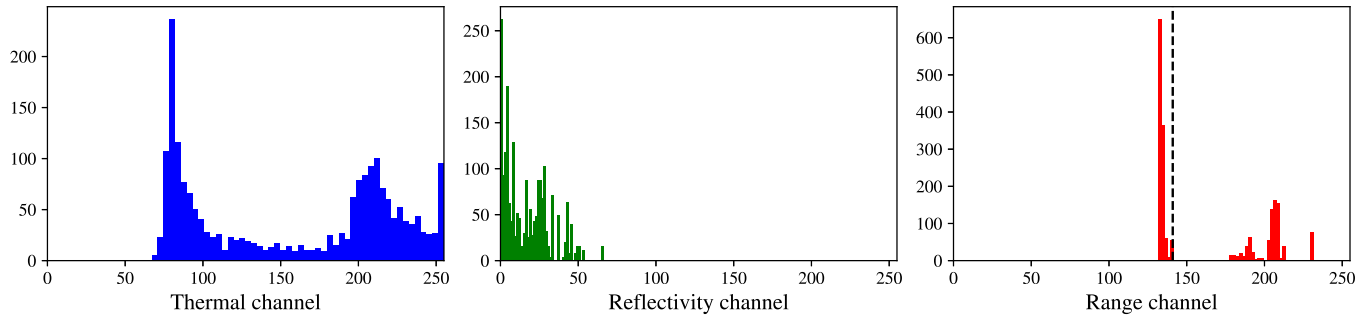


Fig. 19. Histogram of the Region Of Interest for the 3 channels.

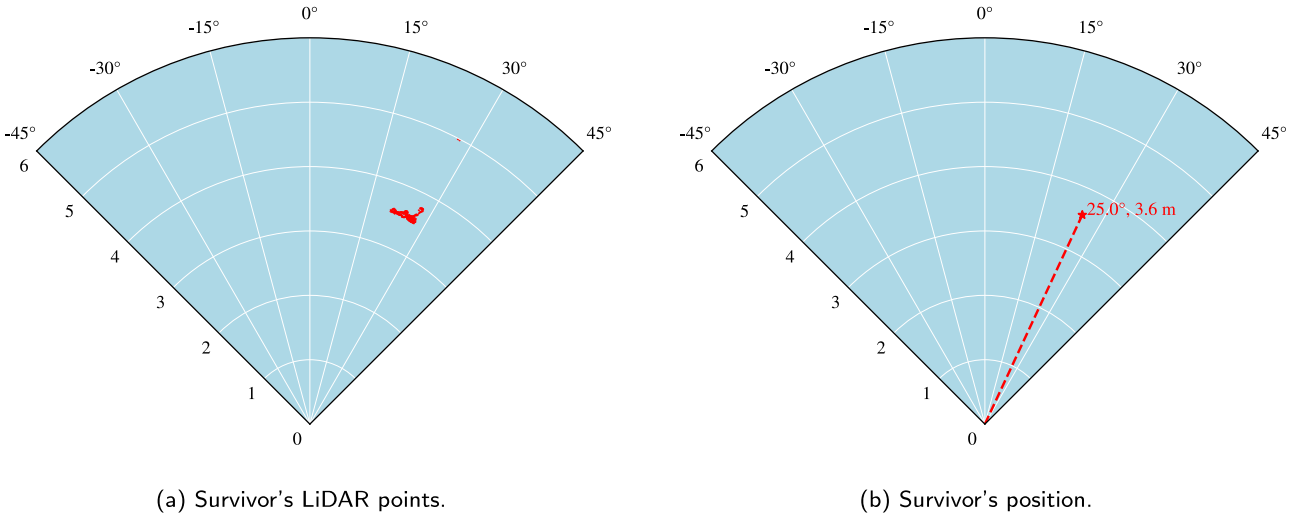


Fig. 20. Survivor detection in the navigation plane.

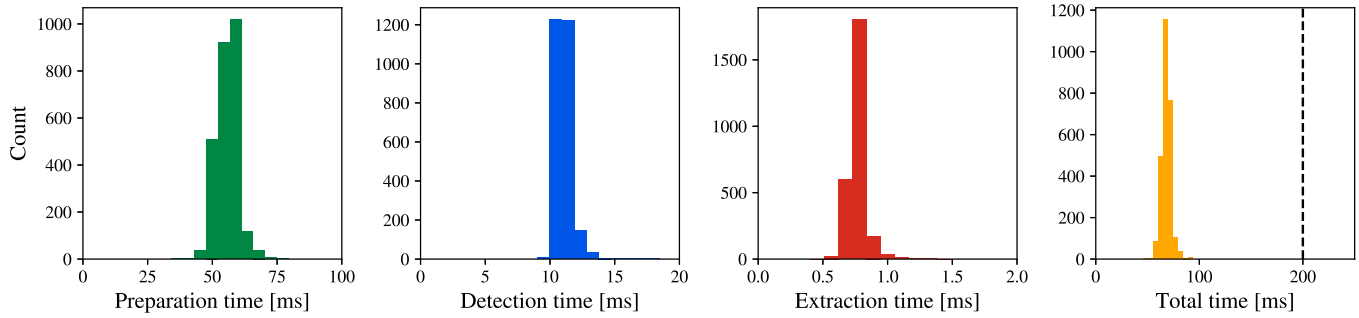


Fig. 21. Pipeline computational times distribution.

Table 7
Complete pipeline computational times.

Time	Mean value [ms]	Standard deviation [ms]
Preprocessing	55.86	4.19
Detection	11.82	0.61
Extraction	0.76	0.07
Total	68.46	4.57

overall performance. Additionally, preprocessing (which dominates the processing time) was not specifically optimized for parallel processing, leaving headroom for real-time implementation on less powerful embedded platforms.

As an additional validation, a general late-fusion pipeline was implemented, combining the YOLO detector trained on thermal-only data with a LiDAR processing branch based on the DBSCAN clustering algorithm, as described by Faggioni et al. (2022a). The overall computational cost of this approach was 188.57 ms, with a standard deviation of 13.68 ms. The early fusion strategy, employing detectors on multi-modal images, demonstrated greater robustness to environmental disturbances in terms of both precision and recall, and exhibited a significantly lower false positive rate as shown in Section 5.1, all while requiring only 36 % of the computational cost.

6. Discussion

The analysis of weather conditions supports the superiority of multi-modal detection compared to single-source detection, providing enhanced extrapolation capacity during training and greater resilience to more challenging conditions than those for which it was trained. The fusion of thermal imaging and LiDAR data enhances the robustness of the detection system, as it is less susceptible to environmental factors.

Furthermore, the multi-modal approach has proven to be less prone to false positives as the severity of atmospheric conditions increases. The validation of detection on a 3-channel multi-source image, constructed from the early fusion of thermal and LiDAR imaging, yielded satisfactory values for the evaluation metrics, which remained consistent throughout the k-fold validation test split. The comprehensive testing conducted across the entire pipeline demonstrates the system's capability to extract information from both sensors and utilize it effectively to provide a precise position estimate for the human-in-water. The computational cost test further validates the system's ability to meet the real-time constraint of 5 Hz, corresponding to the LiDAR acquisition frequency. The system was validated on an off-the-shelf computer, showcasing the scalability and integration potential of the method on a standard onboard processing unit, which is critical for real-time deployment in autonomous vehicles. Finally, the multi-modal early fusion approach saved over 60 % of the computational cost compared to a similar multi-modal approach based on late fusion. Despite the encouraging results, some limitations can be identified. Although the authors made significant efforts to create a dataset closely aligned with real-world scenarios, the data were acquired in a controlled laboratory environment. Additional validation campaigns with data collected in operational conditions are necessary (and are planned for the future) to further validate the pipeline in real-world settings. Finally, given the preliminary nature of the acquired data, the dataset does not yet include ship motion data, which would be critical for integrating the system onto an Autonomous Surface Vehicle.

7. Summary and future work

The ability to quickly locate survivors and provide direct support is a critical aspect of marine disaster management and SAR operations. Conducting these operations using Autonomous Surface Vehicles allows for mission execution without exposing rescuers to risks, while still providing the advantages of having a naval unit deployed on-site. The authors have proposed an effective method that meets the real-time constraint for detecting and estimating the position of a human in water during SAR operations. The choice of sensors and the use of early fusion techniques make the system resilient to adverse marine conditions, which preliminary research has shown to be typical of SAR scenarios. This research aims to provide enabling technologies for the automatic rescue of humans, to preserve human life during dramatic situations, such as maritime disasters.

Despite the promising results, several aspects can be further developed. An experimental acquisition campaign in relevant operational environments is planned to further validate the detection capabilities. The integration of ship motion and position data, obtained through proprioceptive sensing layers, will allow for the translation of the human's position from the sensor reference frame to the Earth-fixed reference frame. Lastly, although the system is prepared, it still needs to be integrated with onboard navigation and anti-collision systems to calculate a correct collision-free route to rescue the survivor.

Declaration of generative AI and AI-assisted technologies in the writing

During the writing of this paper, the authors have made use of generative AI and AI-assisted technologies to improve the language and readability. The authors reviewed and edited as needed the generated text and take full responsibility for the content.

CRediT authorship contribution statement

Filippo Ponzini: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization; **David Van Hamme:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization; **Michele Martelli:** Writing – review & editing, Validation, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was partially funded by European Union – NextGenerationEU. Piano Nazionale di Ripresa e Resilienza, Missione 4 Componente 2 Investimento 1.4 "Potenziamento strutture di ricerca e creazione di "campioni nazionali di R&S" su alcune Key Enabling Technologies". Code CN00000023 – Title: "Sustainable Mobility Center (Centro Nazionale per la Mobilità Sostenibile – CNMS)". However, views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible.

References

Akbar, A.Z., Fatichah, C., Dikairono, R., 2022. Autonomous surface vehicle in search and rescue process of marine casualty using computer vision based victims detection. In: 2022 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM), pp. 1–6. https://doi.org/10.1109/CENIM56801.2022.10037319.

Ancy Micheal, A., Sivaramakrishnan, S., 2024. Human detection and tracking for drone based marine surveillance. In: 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–7. https://doi.org/10.1109/ICCCNT61001.2024.10723860.

Boretti, A., 2024. Unmanned surface vehicles for naval warfare and maritime security. J. Defen. Model. Simulat. 15485129241283056. https://doi.org/10.1177/15485129241283056.

Brandt, P., Munim, Z.H., Chaal, M., Kang, H.S., 2024. Maritime accident risk prediction integrating weather data using machine learning. Transport. Res. Part D: Transp. Environ. 136, 104388. https://doi.org/10.1016/j.trd.2024.104388.

Bryony Gooch, B.D., 2025. North sea: everything we know about oil tanker and cargo vessel collision off Yorkshire coast. Accessed: 2025-04-10. https://www.independent.co.uk/news/uk/home-news/north-sea-oil-tanker-cargo-collision-.

Cheong, S., Jung, W., Lim, Y.S., Park, Y.H., 2024. Thermal-infrared remote-target detection system for maritime rescue using 3-D game-based data augmentation with GAN. IEEE Trans. Geosci. Remote Sens. 62, 1–13. https://doi.org/10.1109/TGRS.2024.3454983.

Clunie, T., DeFilippo, M., Sacarny, M., Robinette, P., 2021. Development of a perception system for an autonomous surface vehicle using monocular camera, LiDAR, and marine radar. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 14112–14119. https://doi.org/10.1109/ICRA48506.2021.9561275.

(EMSA), E. M. S.A., 2024. Annual overview of marine casualties and incidents 2024. Accessed April 9, 2025. Technical Report, European Maritime Safety Agency. https://www.emsa.europa.eu/publications/item/5352-annual-overview-of-marine-casualties-and-incidents-2024.html.

Faggioni, N., Leonardi, N., Ponzini, F., Sebastiani, L., Martelli, M., 2022a. Obstacle detection in real and synthetic harbour scenarios. In: Modelling and Simulation for Autonomous Systems. Springer International Publishing, Cham, pp. 26–38. https://doi.org/10.1007/978-3-030-98260-7_2.

Faggioni, N., Ponzini, F., Martelli, M., 2022b. Multi-obstacle detection and tracking algorithms for the marine environment based on unsupervised learning. Ocean Eng. 266, 113034. https://doi.org/10.1016/j.oceaneng.2022.113034.

Feraru, V.A., Andersen, R.E., Boukas, E., 2020. Towards an autonomous uav-based system to assist search and rescue operations in man overboard incidents. In: 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), pp. 57–64. https://doi.org/10.1109/SSRR50563.2020.9292632.

Gennarelli, G., Noviello, C., Ludeno, G., Esposito, G., Soldovieri, F., Catapano, I., 2022. 24 GHz fmcw mimo radar for marine target localization: a feasibility study. IEEE Access 10, 68240–68256. https://doi.org/10.1109/ACCESS.2022.3186052.

Guard, U.C., 2023. Recreational boating statistics. Online. Accessed April 9, 2025. https://www.uscgboating.org/library/accident-statistics/Recreational-Boating-Statistics-2023-Ch2.pdf.

Helgesen, Ø.K., Vasstein, K., Brekke, E.F., Stahl, A., 2022. Heterogeneous multi-sensor tracking for an autonomous surface vehicle in a littoral environment. Ocean Eng. 252, 111168. https://doi.org/10.1016/j.oceaneng.2022.111168.

Jian, Z., Liang, Z., Li-nan, Z., Nan, L., 2017. The design and research of intelligent search and rescue device based on sonar detection and marine battery. In: 2017 International Conference on Computer Network, Electronic and Automation (ICCNEA), pp. 383–387. https://doi.org/10.1109/ICCNEA.2017.73.

Jocher, G., Qiu, J., Chaurasia, A., 2023. Ultralytics YOLO. https://github.com/ultralytics/ultralytics.

Kang, C.M., Yeh, L.C., Jie, S.Y.R., Pei, T.J., Nugroho, H., 2020. Design of USV for search and rescue in shallow water. In: Intelligent Robotics and Applications. Springer International Publishing, Cham, pp. 351–363.

Katsamenis, I., Protopapadakis, E., Voulodimos, A., Dres, D., Drakoulis, D., 2020. Man overboard event detection from RGB and thermal imagery: possibilities and limitations. In: Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3389189.3397998.

Li, D., Yu, L., Jin, W., Zhang, R., Feng, J., Fu, N., 2021. An improved detection method of human target at sea based on Yolov3. In: 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), pp. 100–103. https://doi.org/10.1109/ICCECE51280.2021.9342056.

Liu, K., Yu, Q., Yuan, Z., Yang, Z., Shu, Y., 2021. A systematic analysis for maritime accidents causation in Chinese coastal waters using machine learning approaches. Ocean Coast. Manage. 213, 105859. https://doi.org/10.1016/j.ocecoaman.2021.105859.

Criminal Court of Livorno, Section I, 1998. Judgment no. 179. Issued on October 31, 1998, Case No. 66/95 RG, No. 542/91 NR.

Lygouras, E., Santavas, N., Taitzoglou, A., Tarchanidis, K., Mitropoulos, A., Gasteratos, A., 2019. Unsupervised human detection with an embedded vision system on a fully autonomous UAV for search and rescue operations. Sensors 19. https://doi.org/10.3390/s19163542.

Mansor, H., Norhisam, M.H., Abidin, Z.Z., Gunawan, T.S., 2021. Autonomous surface vessel for search and rescue operation. Bull. Electr. Eng. Informat. 10, 1701–1708. https://doi.org/10.11591/eei.v10i3.2599.

Martelli, M., Faggioni, N., Ponzini, F., 2022. Detecting and tracking multi-object in real marine environment. In: Proceedings of the International Ship Control Systems Symposium. https://doi.org/10.24868/10707.

Martins, A., Dias, A., Almeida, J., Ferreira, H., Almeida, C., Amaral, G., Machado, D., Sousa, J., Pereira, P., Matos, A., Lobo, V., Silva, E., 2013. Field experiments for marine casualty detection with autonomous surface vehicles. In: 2013 OCEANS - San Diego, pp. 1–5. https://doi.org/10.23919/OCEANS.2013.6741348.

Matos, A., Silva, E., Cruz, N., Alves, J.C., Almeida, D., Pinto, M., Martins, A., Almeida, J., Machado, D., 2013. Development of an unmanned capsule for large-scale maritime search and rescue. In: 2013 OCEANS - San Diego, pp. 1–8. https://doi.org/10.23919/OCEANS.2013.6741340.

Nirgudkar, S., DeFilippo, M., Sacarny, M., Benjamin, M., Robinette, P., 2022. Massmind: Massachusetts marine infrared dataset. https://doi.org/10.1177/02783649231153020.

Odetti, A., Bruzzone, G., Ferretti, R., Aracri, S., Carotenuto, F., Vagnoli, C., Zaldei, A., Scagnetto, I., 2024. Lake environmental data harvester (LED) for alpine lake monitoring with autonomous surface vehicles (ASVS). Remote Sens. (Basel) 16. https://doi.org/10.3390/rs16111998.

Panagiotidis, P., Giannakis, K., Angelopoulos, N., Liapis, A., 2021. Shipping accidents dataset: data-driven directions for assessing accident's impact and improving safety onboard. Data 6. https://doi.org/10.3390/data6120129.

Ponzini, F., Fruzzetti, C., Sabatino, N., 2024. Real-time critical marine infrastructure multi-sensor surveillance via a constrained stochastic coverage algorithm. In: Conference Proceedings of iSCSS. https://doi.org/10.24868/11141.

Ponzini, F., Zaccone, R., Martelli, M., 2023. A multi-sensor indoor tracking system for autonomous marine model-scale vehicles. J. Phys. Conf. Ser. 2618, 012008. https://doi.org/10.1088/1742-6596/2618/1/012008.

Ponzini, F., Zaccone, R., Martelli, M., 2025. Lidar target detection and classification for ship situational awareness: a hybrid learning approach. Appl. Ocean Res. 158, 104552. https://doi.org/10.1016/j.apor.2025.104552.

Rivera Velázquez, J.M., Khoudour, L., Saint Pierre, G., Duthon, P., Liandrat, S., Bernardin, F., Fiss, S., Ivanov, I., Peleg, R., 2022. Analysis of thermal imaging performance under extreme foggy conditions: applications to autonomous driving. J. Imaging 8. https://doi.org/10.3390/jimaging8110306.

Rizk, M., Slim, F., Baghdadi, A., Diguet, J.P., 2023. Towards real-time human detection in maritime environment using embedded deep learning. In: Advances in System-Integrated Intelligence. Springer International Publishing, Cham, pp. 583–593.

Stanislas, L., Dunbabin, M., 2019. Multimodal sensor fusion for robust obstacle detection and classification in the maritime robotx challenge. IEEE J. Ocean. Eng. 44, 343–351. https://doi.org/10.1109/JOE.2018.2868488.

Taipalmaa, J., Raitoharju, J., Queralta, J.P., Westerlund, T., Gabbouj, M., 2024. On automatic person-in-water detection for marine search and rescue operations. IEEE Access 12, 52428–52438. https://doi.org/10.1109/ACCESS.2024.3386640.

Thombre, S., Zhao, Z., Ramm-Schmidt, H., Vallet García, J.M., Malkamäki, T., Nikolskiy, S., Hammarberg, T., Nuortie, H., H. Bhuiyan, M.Z., Särkkä, S., Lehtola, V.V., 2022. Sensors and ai techniques for situational awareness in autonomous ships: a review. IEEE Trans. Intell. Transp. Syst. 23, 64–83. https://doi.org/10.1109/TITS.2020.3023957.

Wang, Y., Liu, W., Liu, J., Sun, C., 2023. Cooperative USV–UAV marine search and rescue with visual navigation and reinforcement learning-based control. ISA Trans. 137, 222–235. https://doi.org/10.1016/j.isatra.2023.01.007.

Zaccone, R., 2024. A dynamic programming approach to the collision avoidance of autonomous ships. Mathematics 12. https://doi.org/10.3390/math12101546.

Zhang, L., Wang, H., Meng, Q., Xie, H., 2019. Ship accident consequences and contributing factors analyses using ship accident investigation reports. Proceed. Instit. Mech. Eng., Part O 233, 35–47. https://doi.org/10.1177/1748006X18768917.

Zhang, Z., 2000. A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22, 1330–1334. https://doi.org/10.1109/34.888718.