OERA Pathway Project
Imaging Sonar Data Automation Feasibility Study
FINAL REPORT

June 13, 2022

Project partners:
SOAR - Sustainable Oceans Applied Research Ltd.[1]
Environmental Research Institute, North Highland College UHI[2]
MarineSitu Inc.[3]

List of authors:
Greg Trowse, P.Eng.[1], Benjamin Williamson, CEng PhD[2], Marilou Jourdain de Thieulloy[2], James Joslin, PhD[3], Mitchell Scott[3]

Executive Summary

Multibeam imaging sonars have application to monitoring fish and marine mammal presence and behaviours in the near field of tidal turbine installations, including evaluating avoidance, evasion, and potential blade strikes. Previous work in the Pathway Program has recommended use of the Tritech Gemini 720is, which demonstrated a high level of utility for visually detecting and tracking targets from vessel and bottom-mounted orientations in tidal flows up to approximately 2.5 m/s in Grand Passage, Bay of Fundy, Nova Scotia.

This project focuses on a comparison of two approaches for automated analysis of Tritech Gemini 720is sonar data: (1) an optical-based deep learning detection approach led by Dr. James Joslin, and (2) an approach based on spatial and temporal filtering for target detection and tracking led by Dr. Benjamin Williamson. The scope of this project was developed based on a practical need to increase efficiency in sonar data assessment, working toward methods that can incorporate reliable automation. The project goal is to advance the development of automated methods for detecting, tracking, and classifying acoustic targets in high energy tidal flows. The results will help inform the Department of Fisheries and Oceans Canada, tidal energy developers, and other stakeholders in the design and implementation of effective monitoring systems for tidal energy projects in the Bay of Fundy and beyond.

The primary datasets for analysis include data from an upward-oriented, seafloor-mounted Gemini collected in Grand Passage in October 2020, and from a downward-oriented, vessel-mounted Gemini from Minas Passage collected by SOAR on September 1, 2021. Both of these datasets focus on artificial targets: a V-Wing glider from Dartmouth Ocean Technologies, a ca. 10 cm diameter basalt rock, and a 0.45 kg lead fishing weight. Additionally, a preliminary analysis of bottom-mounted Gemini data from the FORCE site collected by HTEL Energy in February 2022 is included as a case study on algorithm performance. In all cases, the effective range of the Gemini sonar was found to be 30 m to 40 m depending on the size of the target and environmental conditions (bubbles, sediment, zooplankton, and other acoustic scatterers). Beyond this effective range targets are not easily visible amongst the background noise, although this is site, target and orientation specific, to some extent.

The analysis methodology was developed to evaluate the performance of the two methods based on two key metrics: “precision”, here defined as the portion of all predicted targets which were true targets, and “recall”, also known as the true positive rate, evaluates what portion of targets in a database were found by the algorithm.

Both the optical-based deep learning method and the spatial and temporal filtering method are compared to data based on manual annotations made by a trained technician using Tritech’s proprietary SeaTec data acquisition and processing software. SeaTec’s built-in processing tools for automated object detection and tracking are also evaluated and used in the comparison. The Gemini SeaTec software has proven to be reliable for instrument setup and data collection with a user- friendly interface. The software writes proprietary Gemini .ecd

files, though the raw data can be accessed programmatically for conversion and potential compression into alternate formats for storage and analysis. Although lossless conversion was used in this project, note that any conversion or compression may cause data losses which can significantly affect data analysis.

SeaTec

The recall scores obtained using the SeaTec target detection and tracking module with the V-Wing were greater than 60% and 70% for the Grand Passage and Minas Passage data sets, respectively. However, use of the module also led to false positives (i.e., low precision) indicating a possible need for manual review. Recall was reduced significantly with smaller targets including the rock (0.21) and lead weight (0.10). For biological targets of interest, SeaTec detection could be expected to perform reasonably well for detection of large targets with temporal persistence and high contrast relative to the acoustic intensity of the background (e.g., marine mammals, sharks, and schools of fish). SeaTec was not found to be useful for detection and tracking of the small individual potential fish targets in the case study data set provided by HTEL. This was not an unexpected result, given the reliance of image processing with blob detection-type methods on adequate target contrast and persistence.

Optical-based deep learning detection

For the optical-based deep learning method, the presented work explored the efficacy of one algorithm - RetinaNet - to detect candidate targets in multibeam sonar imagery. This algorithm, which we trained as a binary detector to avoid challenging inter-class differentiation, was developed for and is primarily utilized on optical data. However, as the work shows, the algorithm is capable at detecting target objects in multibeam sonar data at low confidence. The caveat is that precision is lower than ideal (0.56 to 0.76), but recall - particularly frame recall, where target localization is not considered - remains encouragingly high (0.82 to 0.99). This algorithm may offer high enough recall to support autonomous target identification.

However, as with all deep learning approaches, the quality and quantity of input training data is of extreme importance. This work reinforced that naïve utilization of models trained outside of test data constraints offers noticeably worse recall and precision. This is especially true when training data targets are noticeably different in acoustic appearance than testing data targets. However, we found that 1) this decrease in model performance may be less pronounced than optical data, likely due to extreme similarity in the acoustic signature of different targets, and 2) these models exhibit some robustness to site variation unless there is an extreme change in noise characteristics between sites.

While the presented work is encouraging, the authors stress that the adaptation of this (or similar) model to live data requires further study and integration development. Namely, the topics of model robustness, appropriate quality and quantity of data, data ‘cleaning’, model augmentation with in situ data, and inter-algorithm comparison should be considered (including novel methods which are specifically aimed at acoustic data). Additionally, while frame recall is high, the same cannot be said for precision. Lower precision levels are due to deliberately allowing low confidence predictions to occur to minimize the likelihood of missing targets.

Missing targets completely was rare. This was partially by design because, as discussed prior, we consider missing true targets to be significantly more damaging than producing false positives. However, several cases did exist, particularly when considering individual target misses instead of whole-frame misses. For the latter two cases (the presence of false positives with and without true positive detections), this is likely due to setting our confidence threshold to levels lower than typical, resulting in low-confidence predictions occurring. In practice, this issue could be mitigated by 1) raising confidence threshold, at the cost of an increased false negative rate, or; 2) the inclusion of more (high quality) training data, to increase object detection confidence.

Future work should explore if the same level of recall can be achieved but with higher confidence and precision.

Detection and tracking through spatial and temporal filtering

Performance of the “Detection and tracking through spatial and temporal filtering” algorithm has been assessed on non-biological targets and applied to a tidal stream dataset containing biological targets. The algorithm was shown to be successful at automated detection and tracking of targets from a raw dataset, i.e., application to a continuous data stream, providing identification of frames containing targets and tracking movement of targets over time. Tracking of targets was used to increase robustness over single-frame detection to avoid spurious returns, i.e., establishing continuous tracks. The approach provided orders-of-magnitude reduction in data volumes while identifying additional biological tracks that were not annotated in manual review.

Results from the datasets containing non-biological targets showed a recall of >96% for Grand Passage data with a precision of 94%, and a recall from 80 to 100% for Minas Passage data with a low precision due to the rope suspending non-biological targets being identified as a target by the algorithm – given the algorithm was designed to detect targets in the water column distinct from the background, detection of the rope is to be expected and unlikely to be in an eventual tidal stream turbine monitoring application. If, for example, mooring ropes were present in an eventual application, these can be masked to improve precision.

The algorithms were tuned to provide a balance between recall (offering regulator certainty) and precision (providing several orders of magnitude lower volumes of data for review, facilitating continuous long-term monitoring). Trials of the algorithm on the HTEL platform dataset demonstrated encouraging results with 75% true detection of single or pairs of fish (so substantially reduced data volumes for manual review). Further detections were marked for additional manual verification as these presented the characteristics of a potential fish track. Importantly, the algorithm provided greater sensitivity than manual annotations, identifying several fish tracks that were not annotated by the reviewer, while preserving robustness identifying tracks rather than spurious detections of noise.

For improved performance, the algorithm thresholds can be tuned as a function of the dataset characteristics (i.e. presence of interference bands, or returns from the seabed). Adaptive filtering preserves sensitivity of detections across flow speeds and across the swath range of the sonar. Sensitivity of the algorithm can also be increased towards further detection of smaller targets. For a fixed instrument deployment (e.g., sonar on a turbine or seabed platform), these thresholds only need to be adjusted once. The algorithm can then run continuously on large datasets, without the need for human input, providing automated outputs of: target detection time, detected frames numbers, and target coordinates, i.e., a track from which regulator metrics such as encounter rate, evasion, movement and behaviour can be derived. This is particularly useful on large datasets for continuous monitoring, such as collected by the DP platform, where manual annotations are not feasible for continuous monitoring.

Automated detection and tracking was also demonstrated from boat-based deployments, where environmental noise (wakes, seabed returns, etc.) requires greater adaptive temporal and spatial filtering to discriminate targets from background.

Whilst the algorithm was successful at detecting tracks of all target sizes, biological and non-biological, annotated and not-annotated, in different environments and deployment configurations, it should be noted that in the far-range (>30 m) of the sonar, the start or end of the tracks were often missing. This is due to the target intensity decreasing into an increasingly cluttered background. This is a natural limitation of the diverging swath of multibeam echosounders (sonar) and dependent on deployment configuration due to returns from the seabed / sea surface. Hence, the algorithm provided detection and tracking – the original requirement – yet for the greatest measures of animal behaviour, track extension would be beneficial. This can be optimized by swath orientation, and future work will aim to improve detection in the far-range, to extend detected tracks using adaptive sensitivity for track extension with the Kalman filter. This could also be investigated using the Optical-based Deep Learning method trained on the previously detected part of the track.

Use of the detection and tracking algorithm through spatial and temporal filtering enables a rapid detection screening that returns both frames containing targets and coordinates of detected targets within those frames. This enables an automated pre-screening of the data, pointing the user directly towards the frames flagged as containing a potential target, drastically reducing operator input from manual review (precision) while preserving a cautious approach to target detection and tracking in light of regulatory priorities (preserves recall). This tool becomes particularly useful on large datasets or continuous monitoring, where manual review of all data rapidly becomes unfeasible.

Future work

Finally, future work should explore the possibility of combining the deep learning-based detection algorithm with filtering techniques such as the "Detection and tracking through spatial and temporal filtering" approach presented in this work. We see two potential ways these approaches can be combined. In the first, the spatial and temporal filtering algorithms can be used to generate annotations for model training of a deep-learning algorithm. In the second, a deep-learning approach could be used to augment a spatial and temporal filtering algorithm for detection and tracking; for example, where the current spatial and temporal filtering algorithm is able to identify tracks, but the extreme start/end is limited by signal-to-noise at swath periphery, increased and adaptive sensitivity to target detection in a focused area could be feasible with a deep-learning approach utilizing the existing frame-based approach, i.e., to identify the bounding box locations of candidate track extensions.

Transferability of techniques

With widespread use of multibeam imaging sonars to monitor fish and marine mammal presence and behaviour in tidal stream sites, and to provide robust continuous nearfield monitoring around turbines irrespective of visibility or illumination, there is wide applicability to the techniques demonstrated here. The tracking and detection algorithms are tunable for different target characteristics to provide robust detection of different species, and the adaptive filtering approach is robust across hydrodynamic conditions of different sites, including the two locations and deployment configurations demonstrated here, and building on other tracking at the MeyGen and EMEC sites in Scotland (Williamson et al., 2021; Couto et al., 2022). The algorithms have also been shown to be transferable between multibeam imaging sonars of different resolutions, building on and extending previous work with a lower-power lower-resolution multibeam imaging sonar (Williamson et al., 2021). Real time processing capability, and parameterisation of raw datastreams (TB / day) into robust detections and target tracks have been shown, together with demonstrating automated detection sensitivity beyond that of a manual observer. Additional validation data through future demonstration, including across different sites and conditions, will support further transferability of the techniques.

Table of Contents

List of Figures

List of Tables

Acknowledgments

- Funding support from the Net Zero Atlantic (formerly Offshore Energy Research Associated) Pathway Program
- Marine operations support for data collection by Gavin Feiel, Allie Flower, Tristan Guest, and Jillian Morris
- Manual data review and annotations by Allie Flower
- Data review and inspection by Nicholas Petzinna
- Halagonia Tidal Energy Limited for the case study dataset

1 Introduction

Multibeam imaging sonars have application to monitoring fish and marine mammal presence and behaviours in the near field of tidal turbine installations, including evaluating avoidance, evasion, and potential blade strikes.

SOAR conducted field experiments in 2020 as part of the OERA Pathway Program to help evaluate the performance of the Tritech Gemini 720is and Teledyne Blueview M900-2250 multibeam imaging sonars for identifying and tracking discrete targets in high-flow environments. These two imaging sonars were the technologies recommended for testing by the subject matter expert for imaging sonars during the first phase (Global Capability Assessment) of the Pathway Program. Previous work in the Pathway Program (Trowse et al., 2020, 2021) has recommended use of the Tritech Gemini 720is, which demonstrated a high level of utility for visually detecting and tracking targets from vessel and bottom-mounted orientations in tidal flows up to approximately 2.5 m/s in Grand Passage, Bay of Fundy, Nova Scotia. The Tritech Gemini 720is operates at 720 kHz and has a maximum effective sampling range of 30-40 m in high tidal flow environments.

Experiments were conducted in two parts, investigating performance of the sonars when operated from a vessel, with downward orientations, and from a sea floor-mounted frame, with upward orientations. Both experiments were conducted in Grand Passage, Nova Scotia. Details on the objectives, methodology, results, and recommendations are available in Field Assessment of Multi-beam Sonar Performance in Surface Deployments (Trowse et al., 2020), and Field Assessment of Multi-beam Sonar Performance in Bottom Mount Deployments (Trowse et al., 2021).

The reports recommend further testing of multibeam sonars in four focus areas, including:

1. observing fish and other marine animals in locations and seasons (times) with high levels of animal abundance and variety,
2. evaluating the most effective sonar orientations for monitoring the near field of tidal turbines,
3. extending analyses to flow speeds that exceed 3 m/s, and
4. increasing efficiency in data assessment, possibly including reliable automation.

The scope of this project was developed to help address 4), with a **project goal** to advance development of automated methods for target detection, tracking, and classification. The project focuses on evaluating two approaches for automated data analysis: (1) using optical-based deep learning, led by Dr. James Joslin and MarineSitu, and (2) using spatial and temporal filtering methods, led by Dr. Benjamin Williamson. The results will help inform the Department of Fisheries and Oceans Canada, tidal energy developers, and other stakeholders in the design and implementation of effective monitoring systems for tidal energy projects in the Bay of Fundy and beyond.

1.1 Objectives

The project objectives are as follows:

1. Test the capabilities of existing algorithms utilized by Drs. James Joslin and Benjamin Williamson to detect, track, and classify targets from bottom-mounted Gemini data collected in Grand Passage.
2. Optimize above algorithm(s) to detect, track and classify targets from Gemini data collected in Grand Passage.
3. Assess algorithm performance using Gemini data collected at the FORCE tidal demonstration site.
4. Develop recommendations for next steps and tools for advancing automation of target detection, tracking and classification using imaging sonars at the FORCE tidal demonstration site.

1.2 Algorithm availability

The algorithms under development by the Environmental Research Institute, North Highland College UHI (ERI) and MarineSitu are not currently publicly available. The developers are interested in demonstration through application to tidal energy projects and can be contacted by email to discuss opportunities for collaboration.

Benjamin Williamson	benjamin.williamson@uhi.ac.uk
James Joslin	james@marinesitu.com
Greg Trowse	greg.trowse@sustainableoceans.ca

2 Methodology

The methodology was developed to evaluate the performance of 1) optical-based deep learning and 2) spatial and temporal filtering methods for automated analysis of data from the Tritech Gemini 720is multibeam imaging sonar. Both methods are compared to detection and tracking results using the proprietary Tritech Gemini SeaTec software.

The primary datasets for analysis include bottom-mounted Gemini from Grand Passage (as in Trowse et al, 2021) and vessel-mounted Gemini from Minas Passage collected by SOAR on September 1, 2021. Both of these datasets focus on artificial targets consistent with those described in Trowse et al. (2020, 2021). Additionally, preliminary analysis of bottom-mounted Gemini data from the FORCE site collected by Halagonia Tidal Energy Limited (HTEL) in February 2022 is included as a case study on algorithm performance.

2.1 Data Collection

2.1.1 Grand Passage

SOAR worked with Dalhousie Ocean Acoustics Laboratory, Clare Machine Works, and Dasco Equipment to design and build an Autonomous Multibeam Imaging Sonar (AMIS) monitoring system to be deployed on the seafloor. The system included a frame with sonar mounts, the Tritech Gemini and Teledyne Blueview sonars, power supply (three 24 V Deepsea Power and Light SeaBattery Power Modules), subsea data acquisition system (sonar control and data storage with an Intel NUC computer and power conditioning inside a Nortek 500 m depth-rated pressure case), and custom cables for power supply and communication. The frame also carried 140 kg of lead ballast. The AMIS monitoring system is shown in Figure 2.1.1.

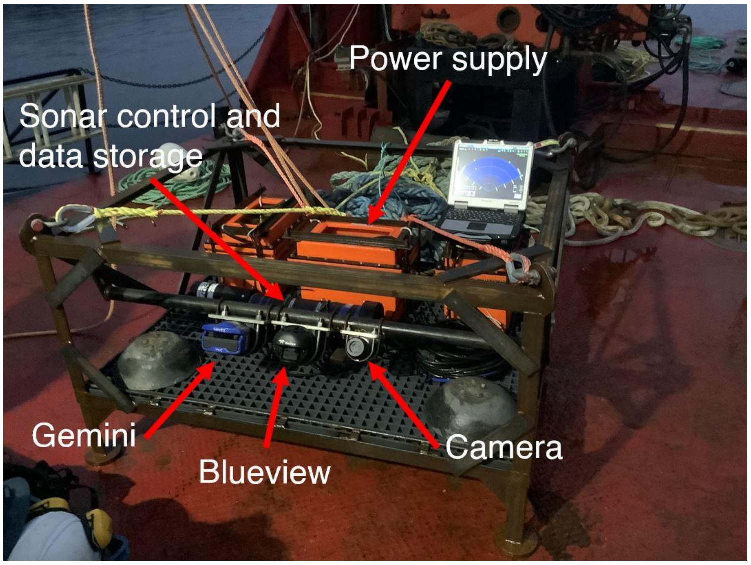


Figure 2.1.1: Autonomous Multibeam Imaging Sonar (AMIS) monitoring system

Two deployments were conducted in Grand Passage in October 2021, hereafter referred to as Deployments 1 and 2. The deployment location is shown in Figure 2.1.2. On both occasions, AMIS was deployed during low water slack and retrieved during high water slack, data being collected during the flood tide. The depth at the deployment location is approximately 25 m at low water, with flow speeds up to approximately 2.5 m/s. A video of the deployment is available at: https://vimeo.com/483103490.

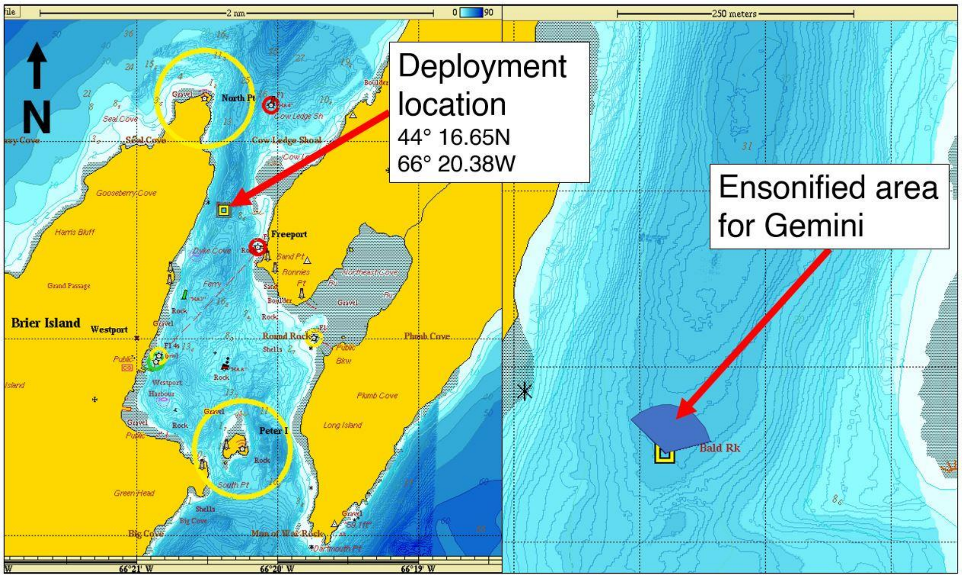


Figure 2.1.2: Grand Passage deployment location

The deployed sonars were oriented such that their ensonified areas were directed downstream, with the instruments' horizontal fields of view oriented across-channel. The configuration was chosen to minimize limitations of the ensonified areas by the sea surface or bottom, while maximizing the horizontal (i.e., downstream) extent over which targets would be visible if drifting downstream at a fixed depth. The horizontal alignment of the instruments was accomplished through use of a ground line and clump weight, which were attached to the AMIS frame. The weight and ground line were lowered first, upstream of the target location for the instrument frame, so that the taut ground line would ensure the correct orientation of the frame when it reached bottom. For both deployments, a diver verified the orientation of the frame and made minor adjustments, and confirmed that no boulders or other obstructions were apparent in the field of view. The diver reported the frame to be sitting well on relatively level ground (less than approximately 5° slope) in both cases.

For Deployment 1, the Gemini was tilted such that the vertical beam width spanned from 5 to 25° above the horizontal plane of the instrument frame. The sampling range of the Gemini was set to 30 m with an associated sampling rate of 13 to 14 Hz. For Deployment 2, the Gemini was tilted such that its ensonified area spanned from 15 to 35° in the vertical. The increase in Gemini tilt for the second deployment was applied due to the presence of consistent returns from the seabed during Deployment 1. The sampling range of the Gemini was set to 50 m with an associated sampling rate of 10 to 11 Hz during Deployment 2.

Schematics of the sonar orientations are provided below, with the plan view shown in Figure 2.1.3 and profile views for the first and second deployments in Figures 2.1.4 and 2.1.5. Example sonograms for the Gemini from Deployment 1 and Deployment 2 are provided in Figures 2.1.6 and 2.1.7.

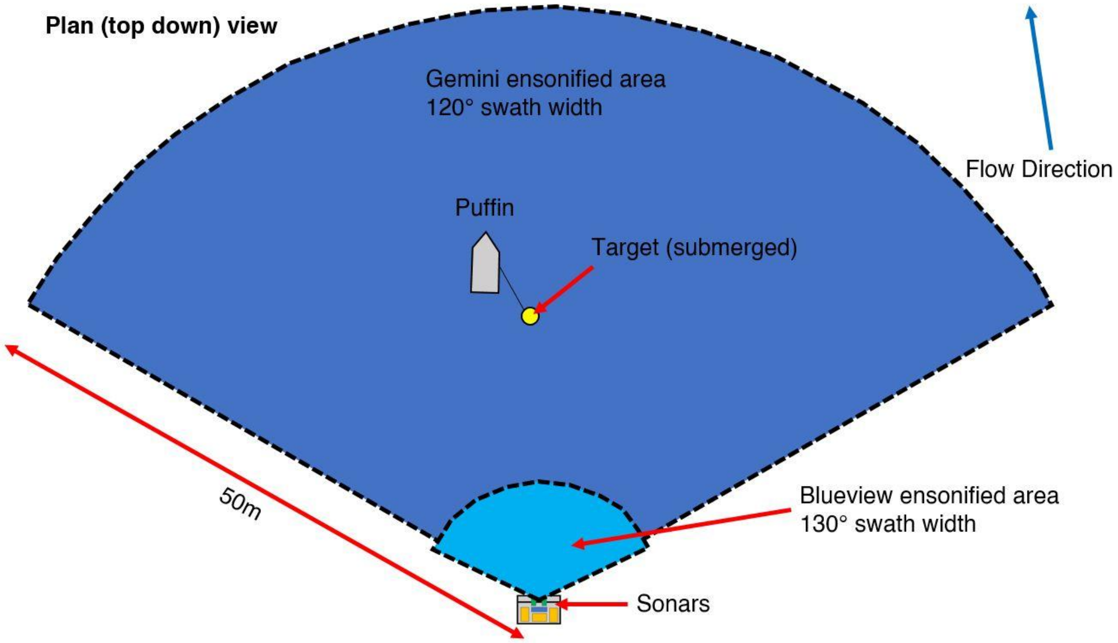


Figure 2.1.3: Grand Passage experiment schematic - plan view

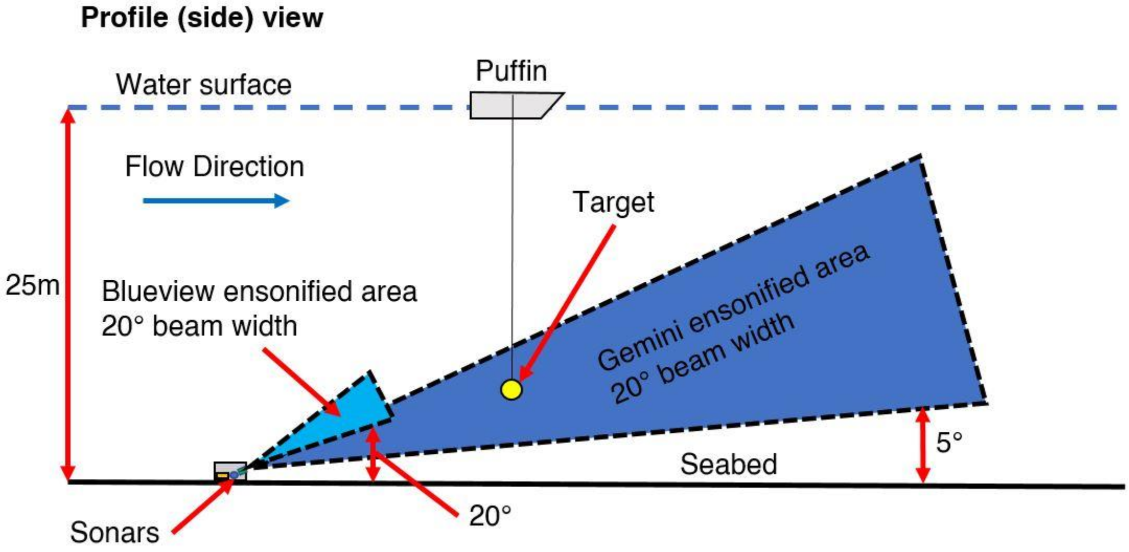


Figure 2.1.4: Grand Passage experiment schematic - profile view - Deployment 1

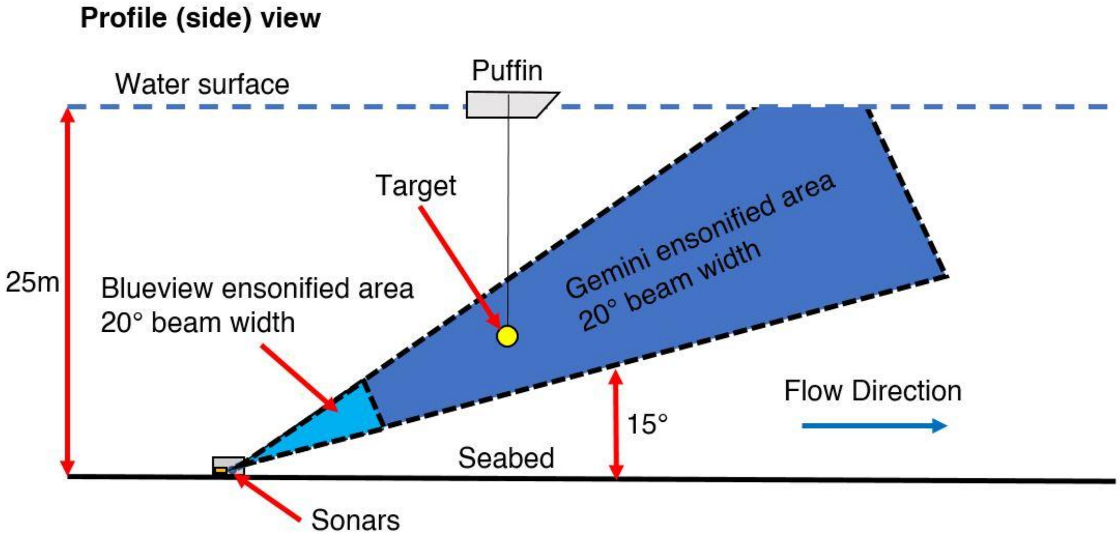


Figure 2.1.5: Grand Passage experiment schematic - profile view - Deployment 2

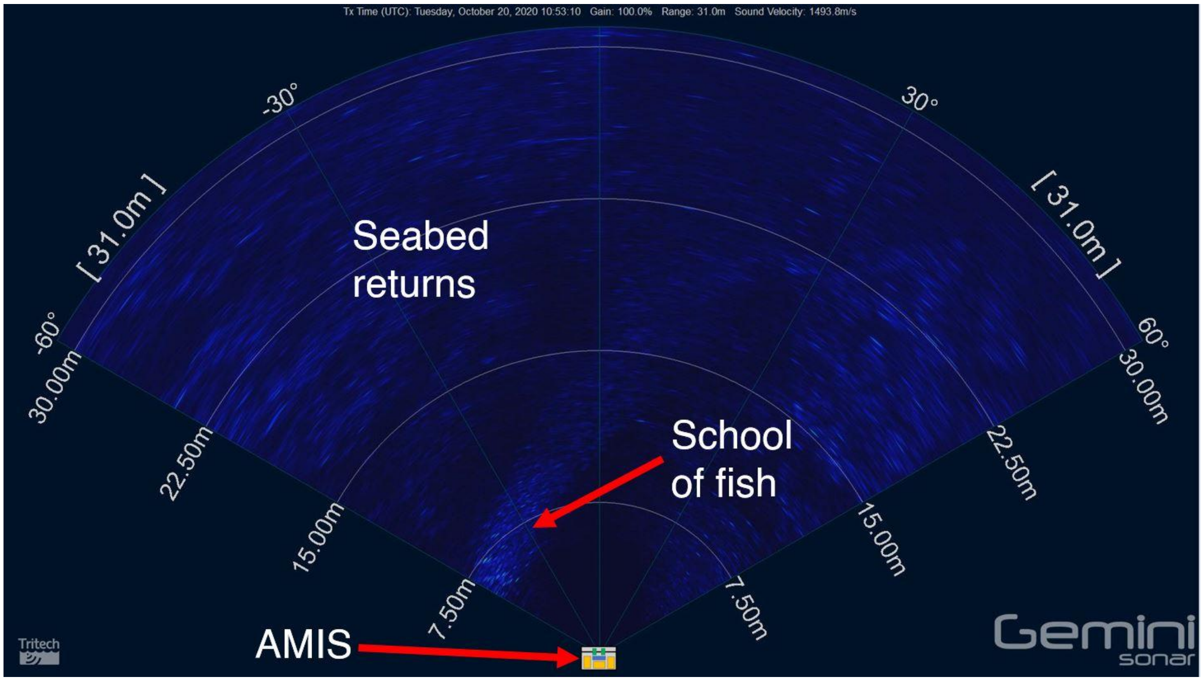


Figure 2.1.6: Grand Passage example Gemini sonogram - Deployment 1

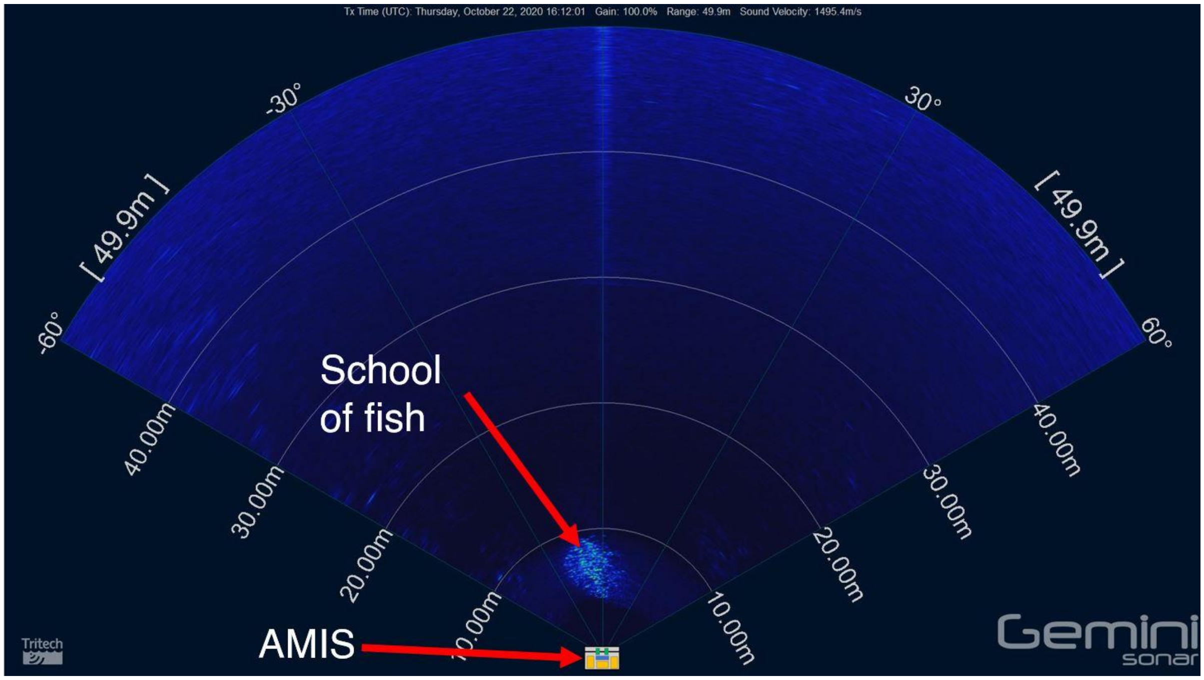


Figure 2.1.7: Grand Passage example Gemini sonogram - Deployment 2

Three targets (shown in Figure 2.1.8) were used during data collection: a 0.45 kg (1 lb.) (9.5 cm long x 3.8 cm max diameter) lead fishing weight (Target 1), approx. 12 cm diameter basalt rock in a lobster bait bag (Target 2), and a V-Wing glider (Target 3) (approx. 52 cm wing tip to tip and 46 cm nose to tail) from Dartmouth Ocean Technologies (DOT). The V-Wing is designed to create downforce and maintain orientation in flow, with approximately (27 kg) 60 lbs. of downforce in 2.5 m/s flow. Quantitative comparative analysis has not been conducted between the acoustic returns of these artificial targets and fish species known to be present in the Bay of Fundy. The target strength of a fish (e.g., with swim bladder) may not be equivalent to the same size target, but these targets were selected to be representative of species commonly found in the Bay of Fundy. Generally speaking, based on experience with the Gemini and observations during data collection, without consideration of life stage and size variation amongst species, we expect acoustic returns from Target 1 to be similar to fish the size of mackerel and herring; Target 2 to be similar to cod, haddock, and salmon; and Target 3 similar to sturgeon, striped bass, halibut, and small sharks. Example sonograms for each target are provided from the work of Trowse et al., 2021 at https://vimeo.com/483141927

Targets were suspended beneath research vessel Puffin (shown in Figure 2.1.8) while drifting through the study area. The Puffin repeatedly travelled to a position upstream from the sonars, then drifted with the tidal flow such that the drift trajectory allowed the targets to pass through the sonars' ensonified areas. The Puffin operated with its dual frequency Raymarine transducer (depth sounder and fish finder) turned off to avoid acoustic interference and collected flow measurements with a RDI 600 kHz ADCP periodically when changing between target types. The ADCP was out of the water during target deployments.

Targets 1, and 2 were suspended from the Puffin using a hand line spool with 200 pound test monofilament fishing line. Target 3 was suspended using 1/4 inch Polysteel fishing line due to the increased downward force, increased cost of the target (reducing risk of loss), and ease of handling. No metal was included in the target suspension system, knots were used to secure the targets with no hooks, shackles, etc. below the water line.

A series of 5 to 15 drifts were conducted for each target, with heights above the seabed that were consecutively increased at 3.6 m (2 fathom) intervals and with minor variations in the drift trajectory to the east and west of the AMIS deployment location. More drifts were conducted for Target 3 due to the higher level of control over depth and horizontal position relative to the Puffin. The AMIS system was fully autonomous, so no live view of data collection was available.

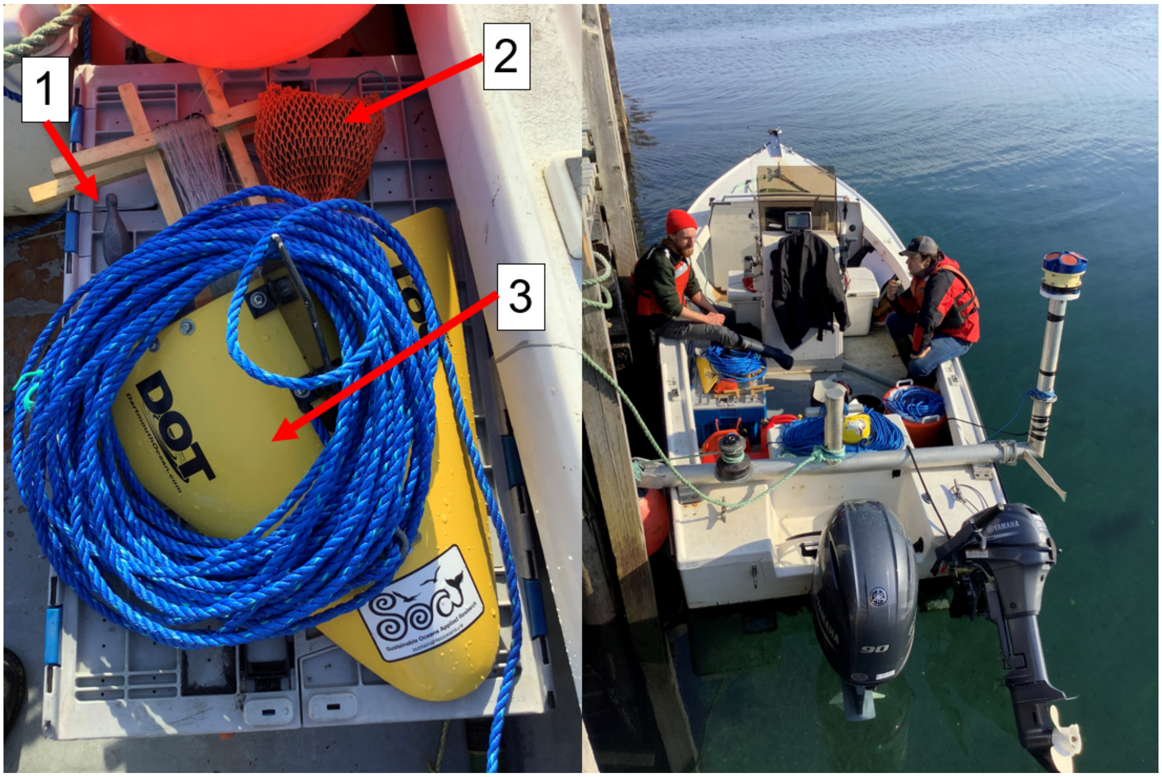


Figure 2.1.8: Targets and research vessel Puffin

2.1.2 Minas Passage

The Minas Passage data set was collected by SOAR at the FORCE site in the Minas Passage on September 1, 2021. The location is shown on Figure 2.1.9.

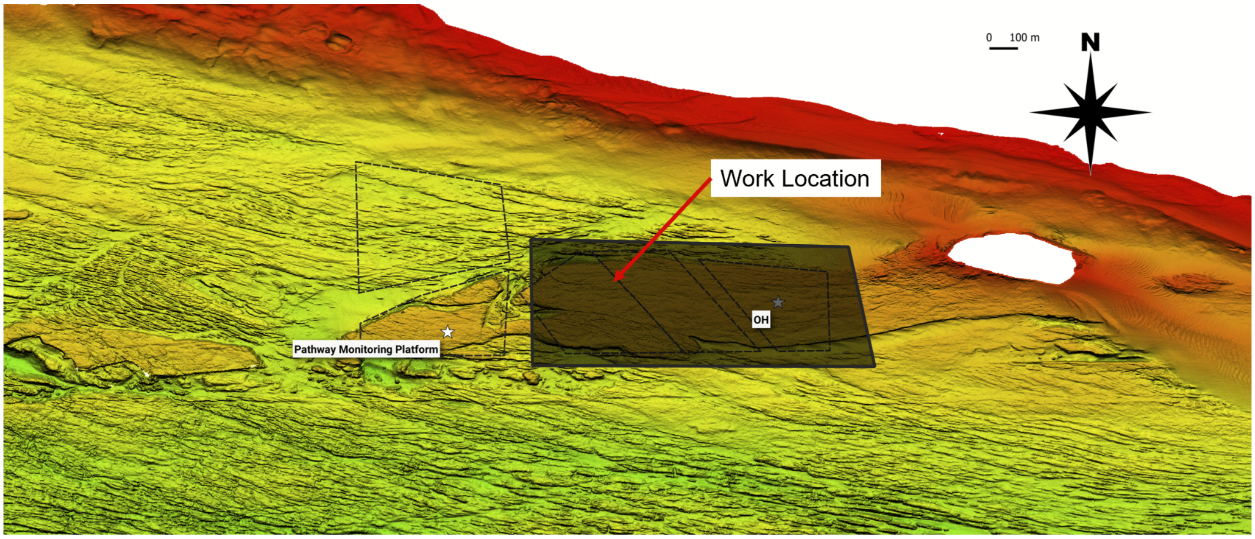


Figure 2.1.9: Minas Passage (FORCE site) data collection location

The Gemini was pole-mounted on research vessel Puffin such that the instrument could be lowered over the side and submerged. The Gemini aimed in the starboard direction away from the vessel, with a downward angle relative to the water surface of approximately 15° (i.e., such that the vertical beam swath spanned from 15° to 35°) to reduce surface returns. Targets consistent with those used in Grand Passage were suspended from a Zodiac. The Puffin held position while the Zodiac deployed the targets approximately 500 m upstream, then drifted through the ensonified area. The targets were observed in real time using the Gemini SeaTec software aboard the Puffin. After the targets passed through to downstream of the ensonified area the Zodiac recovered the target and returned to the upstream position. This procedure was repeated throughout ebb and flood tides, with maximum flow speeds of approximately 4 m/s.

Images of the pole-mounted Gemini are shown in Figure 2.1.10. A target deployment from the Zodiac is shown in Figure 2.1.11.   A schematic of the sonar orientation in plan view shown in Figure 2.1.12 and profile view in Figure 2.1.13. An example sonogram for the Gemini is provided in Figure 2.1.14.

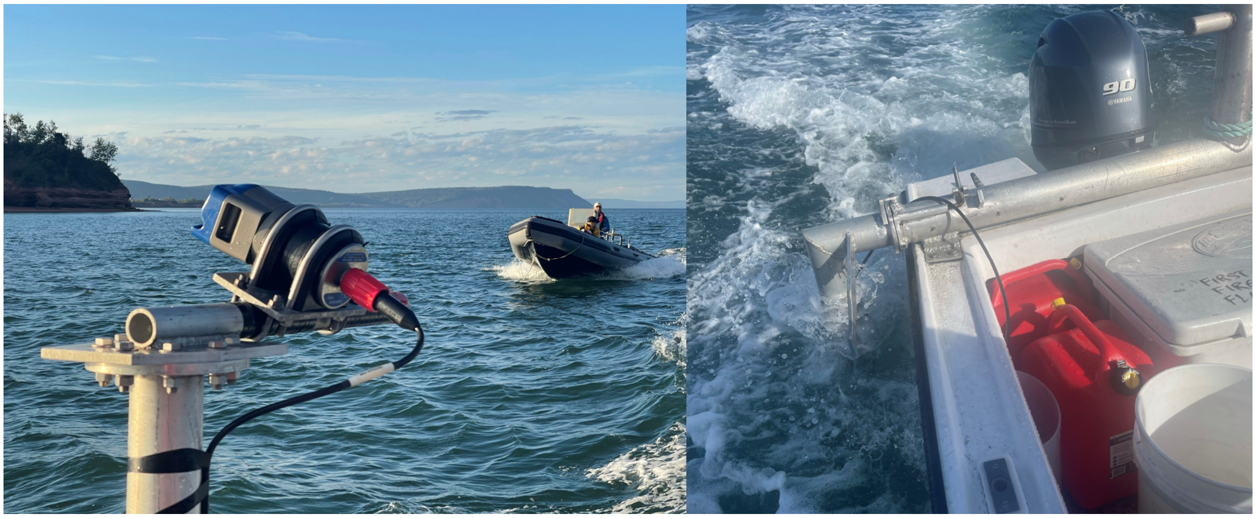


Figure 2.1.10: Pole-mounted Gemini on research vessel Puffin

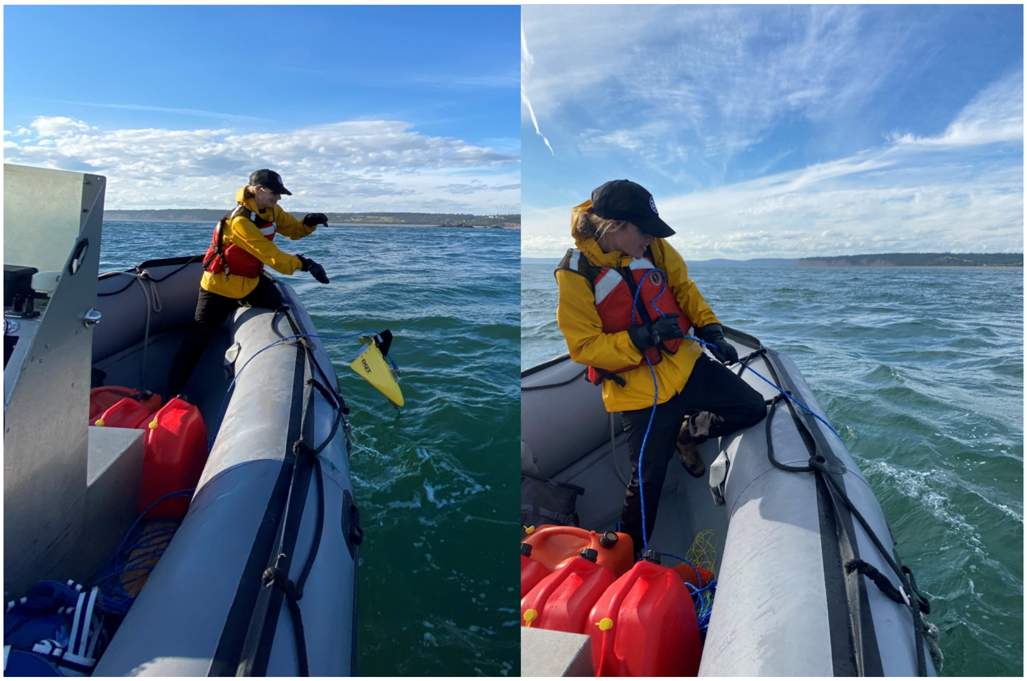


Figure 2.1.11: Target deployment from the Zodiac

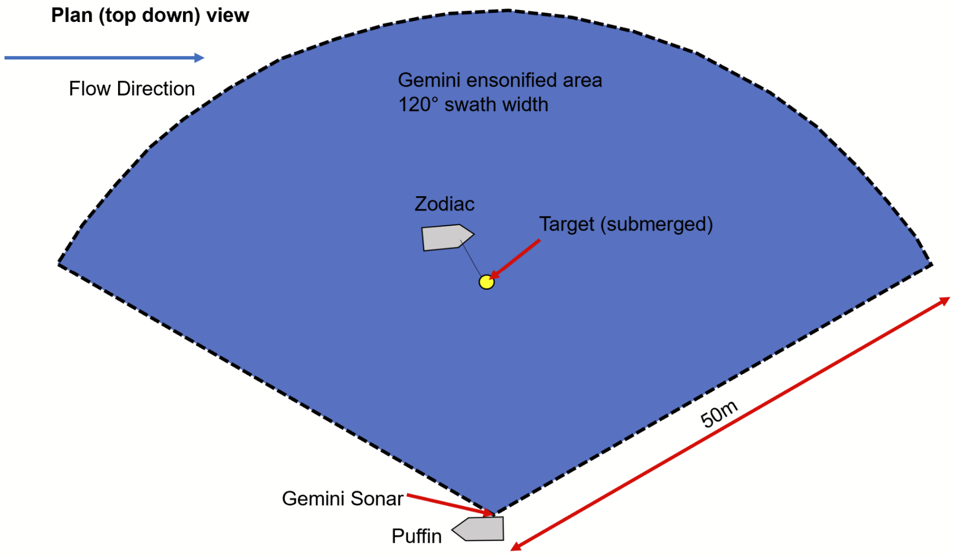


Figure 2.1.12: Minas Passage experiment schematic - plan view

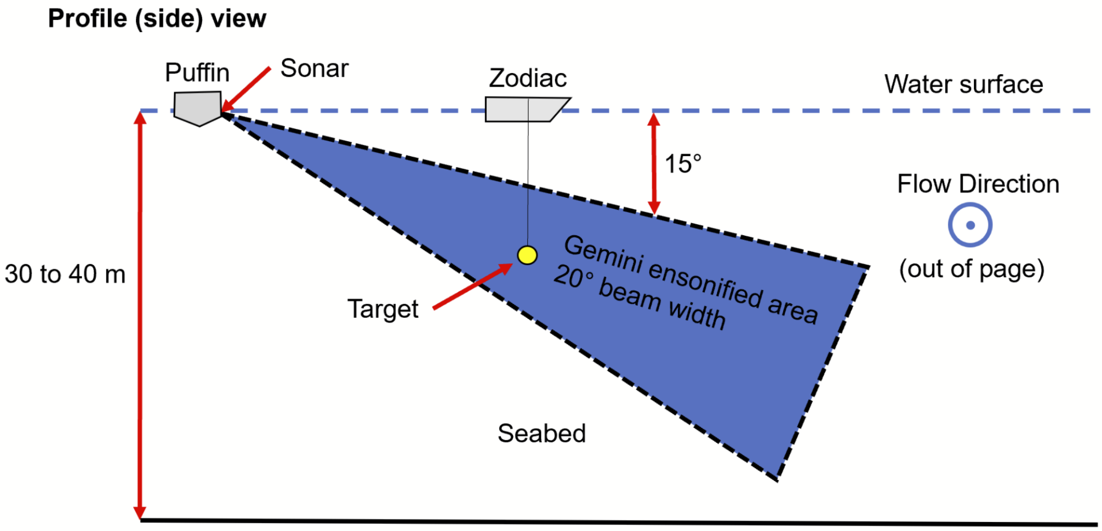
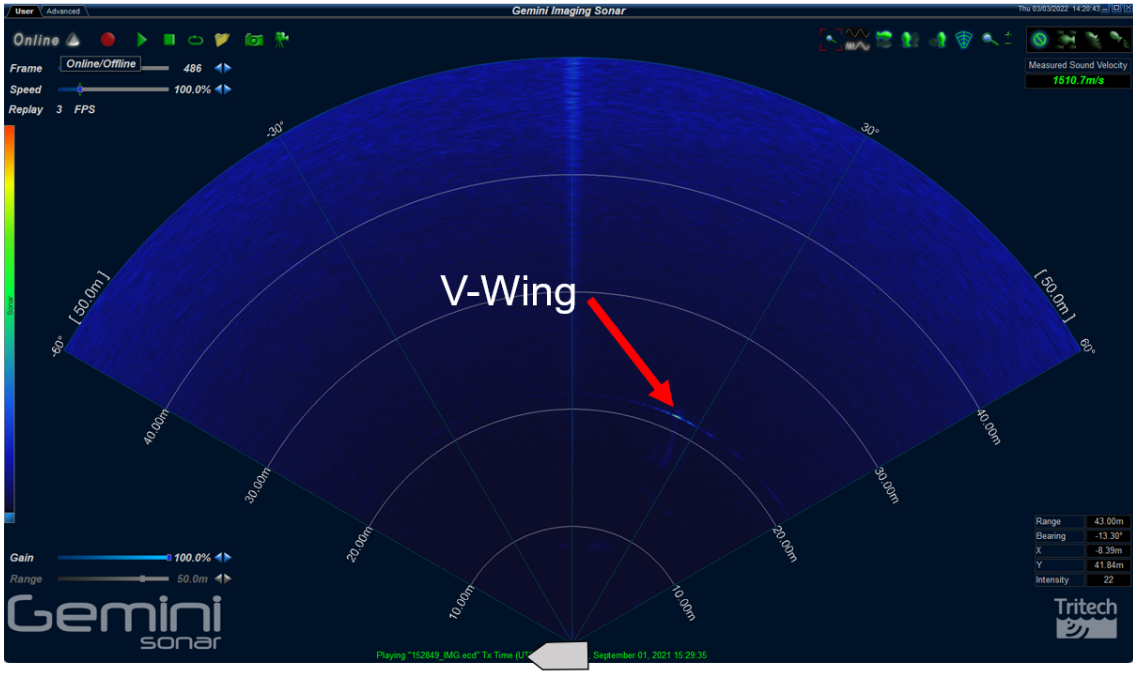


Figure 2.1.13: Minas Passage experiment schematic - profile view



Frame 486

Speed 100.0%

Replay 3 FPS

Measured Sound Velocity 1510.7m/s

Gain 100.0%

Range 50.0m

Range	43.00m
Bearing	-13.30°
X	-8.39m
Y	41.84m
Intensity	22

Playing "152849_IMG.ecd" Tx Time (UT , September 01, 2021 15:29:35

Figure 2.1.14: Minas Passage example Gemini sonogram

2.2 Data Processing

In order to perform data analysis, the .ecd binary files written by the Gemini SeaTec data collection software were converted into suitable formats for the two developed methods for automated analysis, namely the Target Detection using Optical-based Machine Learning methods and the Detection and Tracking through Spatial and Temporal Filtering method.

For the Optical-based Machine Learning method, images (such as .pngs, .jpgs, and .bmps) are required to train and test the model. We therefore first converted .ecd files into images prior to model testing. We chose to use compressed images over raw images for storage reasons, but want to avoid the negative impacts of lossy compression algorithms. Therefore, before training and testing, .ecd files were converted into .png files.

For the data to be used as an input to the Detection and Tracking through filtering algorithm, .ecd binary files were converted into data structures containing 1) a uint8 3D array of size (range bin x beam number x frame) and 2) a parameter structure containing timestamps and instrument settings.

For both of these approaches, data were composed into a polar format grid (Fig 2.1.15) instead of the more commonly utilized cartesian or 'fanned' format (Fig 2.1.14) for displaying sonar data. The motivation was computational efficiency, as conversion to cartesian for display is always possible at a later stage.

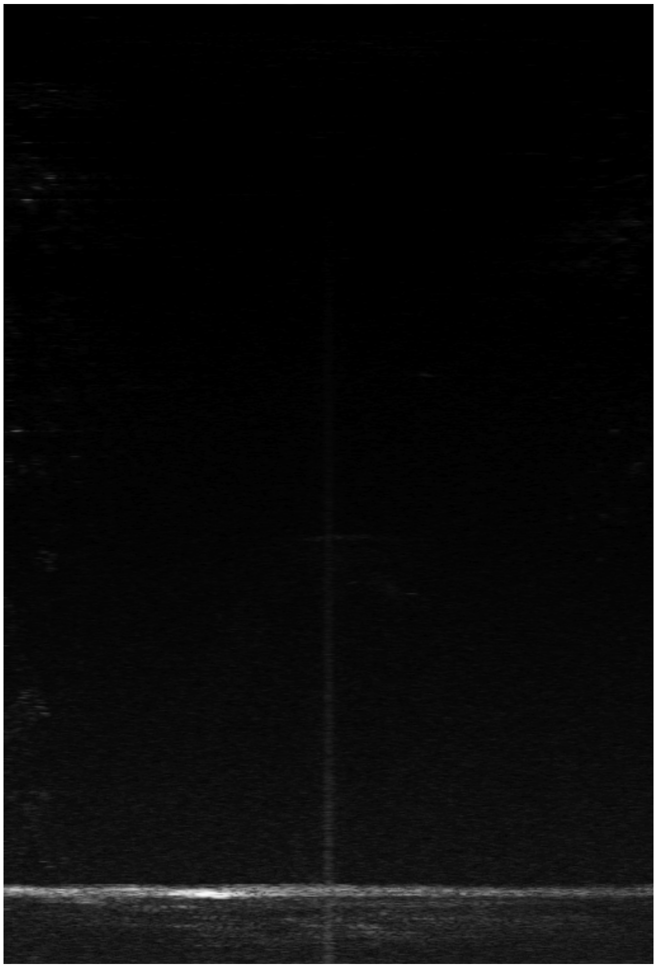


Fig 2.2.1: Example data from Grand Passage, in polar grid format (in contrast to cartesian format in Fig. 2.1.14)

2.3 Data Analysis

During evaluation, we quantified the numbers of true positives, false positives, and false negatives for the different detection methods. This was performed on annotated frames only for the Gemini SeaTec and Optical-based Machine Learning method and on the entire dataset for the Detection and Tracking algorithm (annotated + non annotated frames). A true positive is defined as correctly detecting a target which does exist, a false positive as predicting a target which does not exist, and a false negative as missing a potential target. Note that the definitions of correct and incorrect detection is method-specific, and is described by each of the methods in the following sections.

By gathering these classifications, we can define the precision and recall of an approach. These two metrics, which are commonly utilized to present the accuracy of detection-like algorithms, help indicate how well a tuned algorithm can detect targets in an environment and the ratio of predictions which are incorrect. Specifically, precision indicates the portion of all predicted targets which were true targets, i.e.:

$$P = \frac{tp}{tp + fp}, \qquad \text{Eqn. 1}$$

For precision P, true positive tp, and false positive fp. Recall, which is also known as the true positive rate, evaluates what portion of targets in a database where found by the algorithm:

$$R = \frac{tp}{tp + fn}, \qquad \text{Eqn. 2}$$

for recall R and false negative fn.

2.3.1 Gemini SeaTec

The Gemini SeaTec software developed by Tritech was used for:

- manual data analysis, including annotating target detection and tracks for training and testing automated analysis, and
- testing of the Gemini SeaTec target detection feature for comparison to the a) Target Detection using Optical-based Machine Learning and b) the Detection and Tracking through Spatial and Temporal Filtering methods.

Annotations included the file name, time, ping number, a bounding box for the target in Cartesian coordinates (lower X, lower Y, upper X, and upper Y), target description, and general notes. Approximately 550 annotations were logged into a Google Sheet for the Grand Passage data set and 1,000 for the Minas Passage data set. Google Sheets were used to facilitate sharing amongst the project team, and easily export as .csv format for programmatic training and evaluation of approaches for automated analysis.

SeaTec uses an image processing blob detection method that detects regions in a digital image that differ in properties compared to surrounding regions. Tritech provided *the following* description of the approach.

SeaTec uses differences in intensity to detect and track the targets. This includes three steps i) remove static targets, ii) locate moving targets, and iii) verify moving targets. The image processing flow is shown in Figure 2.3.1.

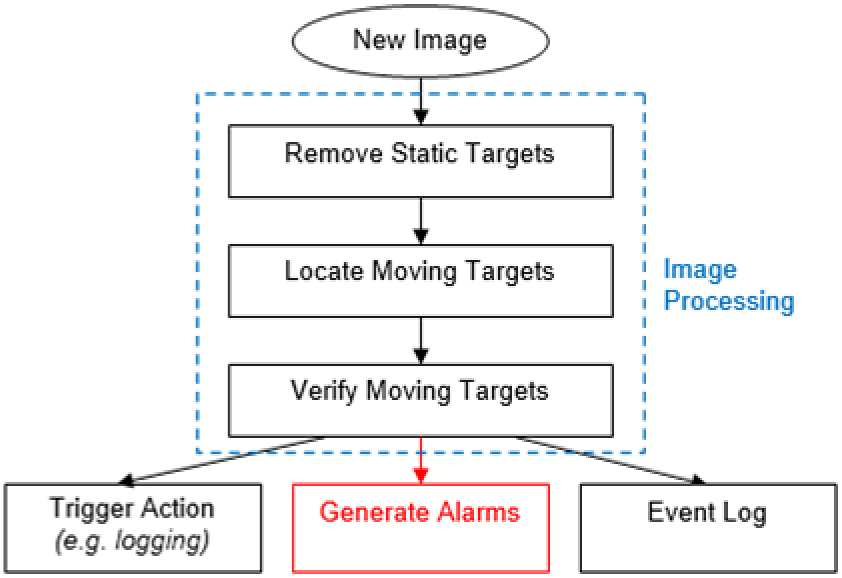


Figure 2.3.1: SeaTec image processing flow.

The first step of the process uses accumulation and subtraction buffers to remove persistent targets. The accumulation buffer is continually updated using scaled intensity values from each image. Weighted intensity values from the current image are then subtracted from the accumulation buffer to remove static and semi-static regions of high intensity.

After removing the static targets, the entire image is analyzed for pixels that exceed a detection intensity threshold. When a pixel is detected with higher intensity, a flood fill is performed from that point, using a lower flood intensity threshold. This determines the extent of the target. The number of points within the target that exceed the detection intensity threshold are also counted. The default flood intensity threshold is 5%.

After that several tests are performed to discriminate between a valid target and other targets or noise in the sonar image.

1. *New targets are checked for overlap with targets identified in the previous sonar image, to remove glitches that are present in a single sonar image only. If the target does not overlap a previous target, then that target is classed as a transient target and is not considered valid.*
2. *If the number of points within the target that exceed the detection intensity threshold is less than a certain percentage of the total number of points in the target, then that target is classed as a weak target and is not considered valid. The default percentage of points that must exceed the detection intensity threshold is 4%.*
3. *If the dimensions of the target do not exceed minimums specified by the user, then the target is classed as a small target and is not considered valid.*
4. *If the dimensions of the target exceed maximums, then the target is classed as a large target and is not considered valid.*

5. *If the target is not within minimum and maximum ranges, then the target is classed as an out of range target and is not considered valid.*
6. *If all the above tests pass, then the target is considered valid.*

The SeaTec software was tested using the default settings for detection and tracking sensitivity. Minimum target size was set to 0.01 m and maximum to 3 m based on the size of targets used in our data collection.

For each annotated ping we recorded whether SeaTec detected the target (1 for yes, 0 for no) as well as the number of additional targets which SeaTec classifies as either probable, possible, or potential. For example, if the target was detected along with an additional 4 the resulting recall is equal to 1 and precision is equal to 0.2.

2.3.2 Target Detection using Optical-based Machine Learning methods

2.3.2.1 Method overview

MarineSitu is currently developing custom software for environmental monitoring of marine energy sites using machine learning methods to autonomously detect targets in optical and sonar imagery. For this study, the first approach explored target identification using common machine learning detector algorithms. While these algorithms are primarily developed for and utilized with optical camera systems, several approaches have explored their ability to be trained and tested on acoustic data (Valdenegro-Toro 2019, Neupane 2020, Singh 2021). In the presented study, we aimed to quantify the performance of these algorithms on multibeam sonar data for the detection of rare target events.

What was *not* of interest, however, was rigorous inter-algorithm comparison. That is, we did not attempt to quantify that one algorithm would, in general, perform better when tasked with environmental monitoring. Instead, we focused on training and tuning an individual algorithm to determine if it could be utilized for acoustic data and if so, with what accuracy. Additionally, we favoured algorithms with an already existing codebase, which could be easily integratable into a software stack and run in real-time.

Our method uses these algorithms for *binary* classification. We train the algorithm to recognize targets, not distinguish between targets (e.g., the utilized method could, in theory, be trained to detect fish, but not distinguish between fish *species*).

2.3.2.2 Algorithm overview

Our work therefore focuses primarily on one algorithm for target identification: *RetinaNet* (Lin 2017a). This fully supervised approach, from Facebook AI Research (FAIR), requires class labels at train time, and predicts both class identification and the location of a target in an image by producing a bounding box. Its architecture is shown in Fig 2.3.2.

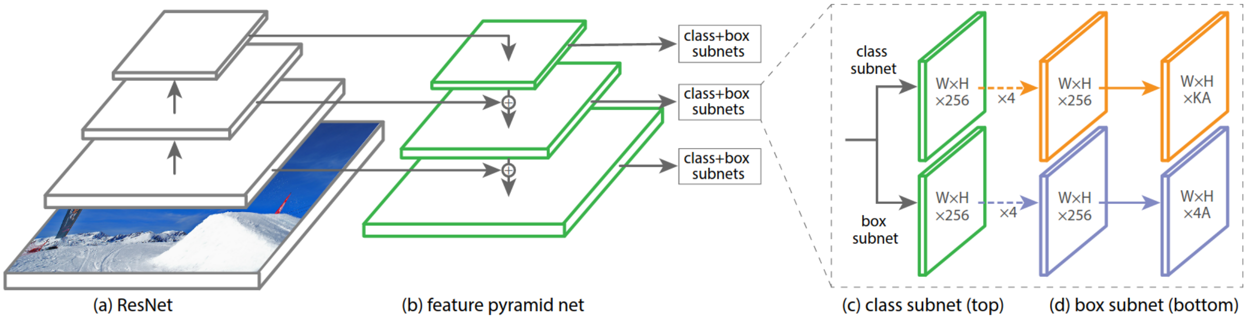


Fig 2.3.2: RetinaNet architecture, presented by (Lin 2017a). Original text: The one-stage RetinaNet network architecture uses a Feature Pyramid Network (FPN) (Lin 2017b) backbone on top of a feedforward ResNet architecture (Huang 2017)(a) to generate a rich, multi-scale convolutional feature pyramid (b). To this backbone RetinaNet attaches two subnetworks, one for classifying anchor boxes (c) and one for regressing from anchor boxes to ground-truth object boxes (d). The network design is intentionally simple, which enables this work to focus on a novel focal loss function that eliminates the accuracy gap between our one-stage detector and state-of-the-art two-stage detectors like Faster R-CNN with FPN (Lin 2017b) while running at faster speeds.

Similar to several other optical object detectors, RetinaNet is a one-stage algorithm, which produces both a bounding box localization and the identification of targets, if present. A novel contribution of their work is the utilization of focal loss, an improved loss quantification to address the *class imbalance* problem, which is caused by the majority of images (and the majority of image locations) in a dataset containing no targets of interest. Given that class imbalance may be a particularly bad problem during environmental monitoring - as the majority of images contain no targets - we believe this algorithm exhibits promising attributes for utilization in rare target detection when compared to algorithms which do not directly address this problem.

This algorithm is trained by splitting curated data and its associated labels into training and testing partitions, similar to other fully-supervised approaches. This common procedure verifies that evaluation does not take place on a model which has been “overfit” to a particular dataset. Example *inferences* (i.e., algorithm predictions) are shown in Figs 2.3.3 and 2.3.4. In Fig 2.3.4, a false positive detection is also shown, near an interference band.

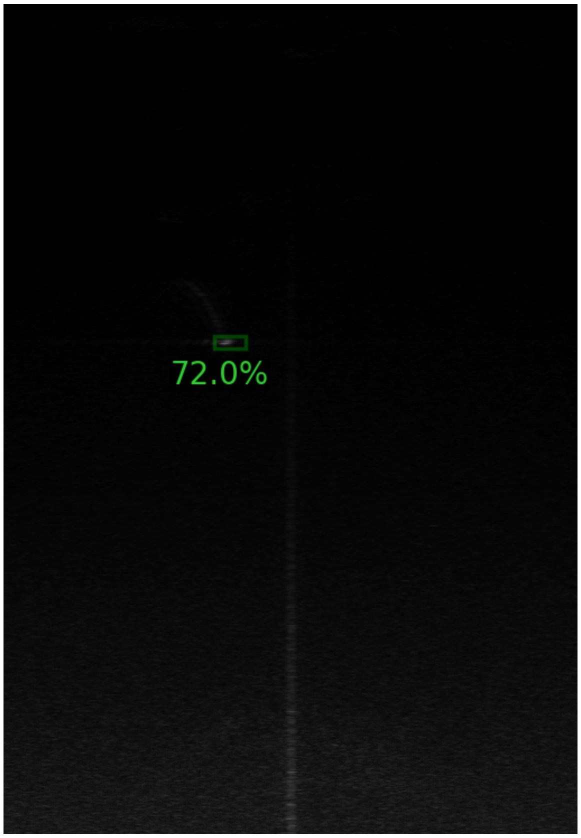


Fig 2.3.3: Detection of a target at the Minas Passage site using RetinaNet. The 72% indicated near the bounding box is the confidence *the model assigns to the prediction.*

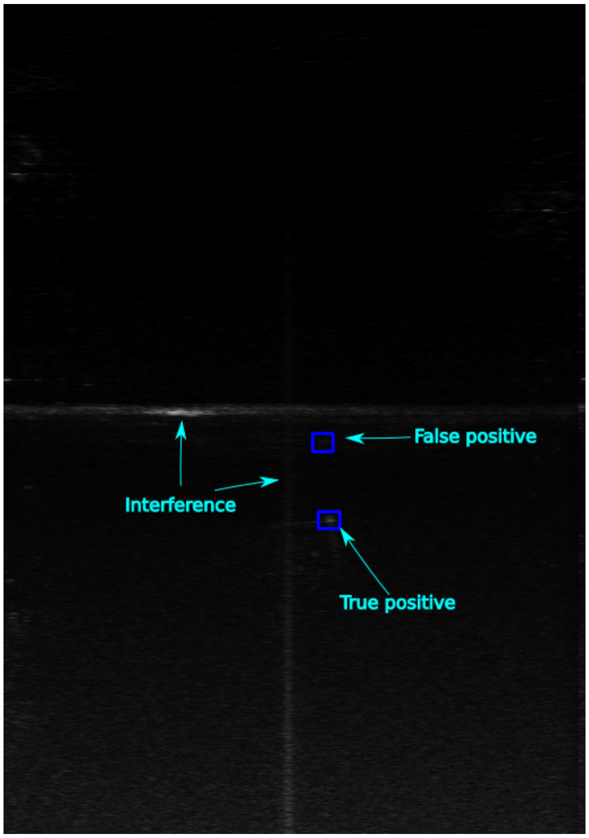


Fig 2.3.4: Detection of a target at the Grand Passage site using RetinaNet. True and false positive shown

2.3.2.3 Evaluation

Evaluation of this method utilizes precision and recall, as discussed in Section 2.3. As with training, evaluation requires annotated frames. However, given that our chosen algorithm produces bounding boxes instead of whole-image classification, an additional step is required to determine if a predicted bounding box is a correct classification. For that step, *Intersection over Union (IoU)* is utilized to evaluate if a predicted bounding box sufficiently overlaps with a ground truth bounding box. As shown in Fig. 2.3.5, IoU is a simple metric where the overlapping area of a ground truth bounding box and its detected bounding box is compared to the combined areas of these objects.

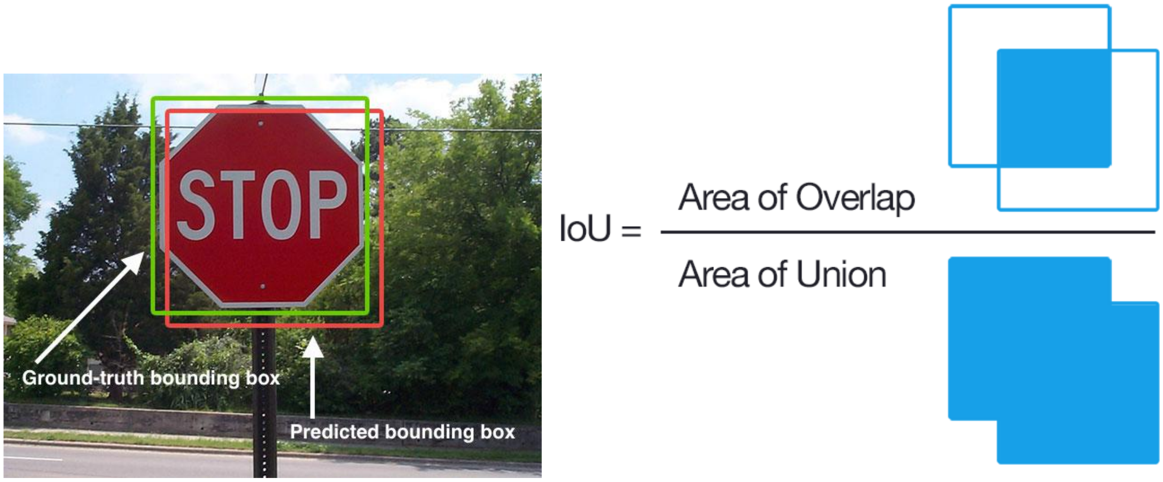


Fig 2.3.5: Intersection over Union (IoU) example. In this case, the ground truth bounding box (green square) is compared to a predicted bounding box (red square), and IoU is the ratio of the area of overlap over the area of union. Photo credit: Adrian Rosebrock, Wikipedia Commons (Rosebrock 2016a, Rosebrock 2016b)

A detected target is considered a true positive if IoU is above a specified 'IoU threshold', and a false positive if it is not. Note in the case of no overlapping bounding boxes - resulting in both the numerator and denominator being zero - IoU is defined as zero.

Typically, this IoU threshold is also utilized to define false negatives. However, for the presented case, we consider a situation where this IoU threshold is completely relaxed for false negative detection. That is, in this scenario, *any* prediction in a frame which contains a target will result in a true positive detection. The motivation here is to not overly 'penalize' the model during evaluation when it informs that a target was present in a frame, even if the true target detection was in a different image location. The accuracy of the predicted bounding boxes is described by the precision metric. Further, manual annotation of targets was not perfect, resulting in some true positive detections not being recognized by the evaluation. **Note:** This approach - defining recall based on targets present in a frame even if bounding box overlap is low or nonexistent - is substantially different from what is typically utilized for image detector evaluations, and is likely only a useful metric for rare-target event prediction like the presented scenario. Given this atypical analysis metric, we also report the more typical metric for recall, where false negatives are recorded if the prediction does not have enough overlap with ground truth to satisfy the IoU threshold.

We denote these two recall cases as: *frame recall* (in which any detection in a frame will not cause a false negative classification) and *IoU recall* (where IoU threshold must be satisfied to avoid a false negative classification). An example of this case is shown in Fig 2.3.6 where a poorly predicted bounding box is shown.

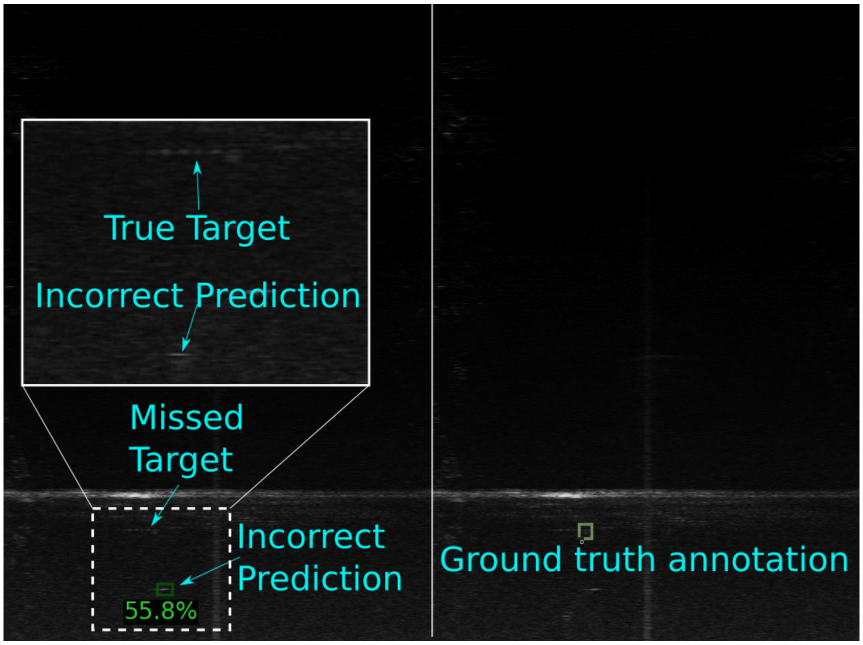


Fig 2.3.6. *Example failure case #1 at Grand Passage. Right: Ground truth target. Left: Predicted target, which is not in the correct location. As an incorrect target is found, and a false positive is recorded. If using IoU recall (sec 2.3.2.3), a false negative is also recorded since the pixel location of the predicted target is incorrect . If using frame recall, a false negative is* not *recorded, as a detection occurred, just in the wrong location.*

The selection of IoU threshold is critical, as its value can have substantial consequences on the evaluation of model performance. An IoU set too low will cause poorly predicted boxes to be considered true positives, resulting in our evaluation method being overly confident in the model's precision. Conversely, an IoU threshold set too high will result in evaluation which does not consider an overlap to be a true positive detection. For example, in Fig 2.3.3, an IoU threshold set too high will result in a false negative detection when using IoU recall.

Given that the studied application is the detection of rare targets in an environment - a situation where the majority of image frames and areas do not contain a target - we utilize a lower than typical IoU threshold. This low IoU threshold value will result in low false negative levels (e.g., not many true targets are missed) when evaluating on IoU recall, at the expense of higher false positives (e.g., some bounding boxes evaluated as correct are actually incorrect). This trade-off is justifiable for the presented scenario, as an algorithm which misses targets is significantly worse than one which produces too many false positives. Note that this is different from many typical object detection cases (such as web-based search e.g., Google Images), high precision is the primary performance driver. In the presented analysis, we utilize IoU threshold of 0.3 and 0.4.

In addition to IoU threshold, the chosen detection methods must also specify a *confidence* threshold prior to precision and recall analysis. Detection algorithms will often predict many targets in an image. However, many of these detections are very low confidence (e.g., < 10%). The majority of these low confidence detections should be ignored, with only high-confidence predictions considered in further processing. For example, in Fig 2.2.3, an object is detected with a confidence of 72%; a relatively high-confidence prediction for this work.

Following the same logic guiding IoU threshold, confidence threshold is set to lower levels than many typical applications. Again, this is motivated by the belief that false negatives are much worse than false positives, and we favour approaches which keep the number of missed targets to a minimum. In this analysis, the confidence threshold is set at a level of 0.3-0.4.

2.3.3 Detection and tracking through spatial and temporal filtering

ERI has previously developed and applied algorithms for automated detection and tracking of animal movement and behaviour using multibeam sonar in tidal stream energy sites (Williamson et al. 2021).

As part of this project, ERI developed and implemented a new algorithm with the following aims:

1. Across a whole dataset, identify the frames containing a potential target;
2. Within the identified frames, save target centroid coordinates;
3. Build target track(s).

Performance of the algorithm is assessed against a baseline of manual target annotation. Results of the comparison are displayed in Section 3.3.

Algorithm development was performed in MATLAB 2021b but can be compiled and run standalone for future applications.

The target detection and tracking algorithm contains the following steps, which are detailed in the next subsections:

1. Detection of candidate targets
2. Temporal filtering
3. Target clustering and track reconstruction using Kalman Filter
4. Results export

2.3.3.1 Detection of candidate targets

Detection of candidate targets is a per-frame process, based on the difference between image background, interference and target properties in time and space.

- Background intensity is assumed to vary sequentially across range and acoustic beams, and be mostly consistent over time.

- Acoustic interference has been classified into two types: acoustic interference bands, that are sporadic and contain high intensity across an entire range cell or beam, and acoustic interference reflections, whose high intensity repeats at a constant location across multiple frames (temporally persistent).
- Potential targets are defined as isolated elements made of a small group of pixels, whose intensity is higher than the mean background intensity across range, beam and time using adaptive filtering which preserves sensitivity despite temporally and spatially varying noise.

Based on the above definitions, the detection process is performed for each individual image frame. Image frame numbers containing candidate targets are saved along with the target coordinates.

Detection is performed in polar coordinates for computational efficiency and coordinates of candidate targets are subsequently transformed into Cartesian space.

Figures 2.3.7a and 2.3.7b show an example of a frame containing background noise, interference, and target pre and post detection, with pixels containing detected target candidates are shown in red. Figure 2.3.8 shows the time superposition of the automatically detected targets in Cartesian coordinates. The target detection step is conservative, seeking to preserve all possible candidate targets (including false positives) which are removed during the next steps, rather than the converse strict approach which would minimize false positives, but also risk false negatives which has far greater implications (missed detections).

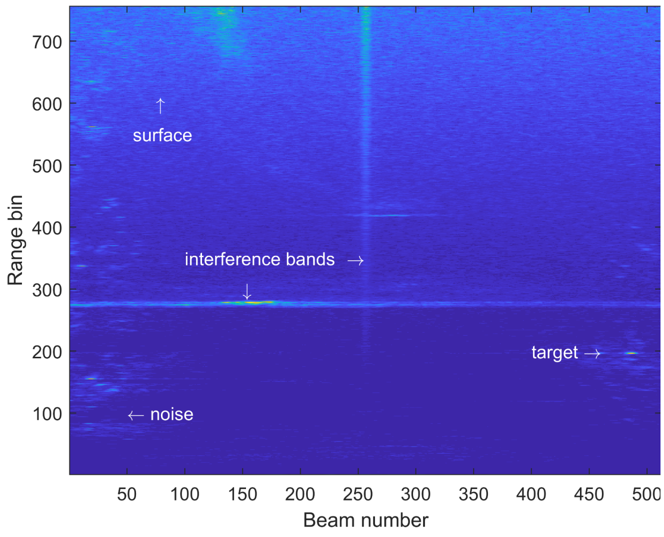


Figure 2.3.7a: Image frame in polar containing post-detection, background noise, interference and target.

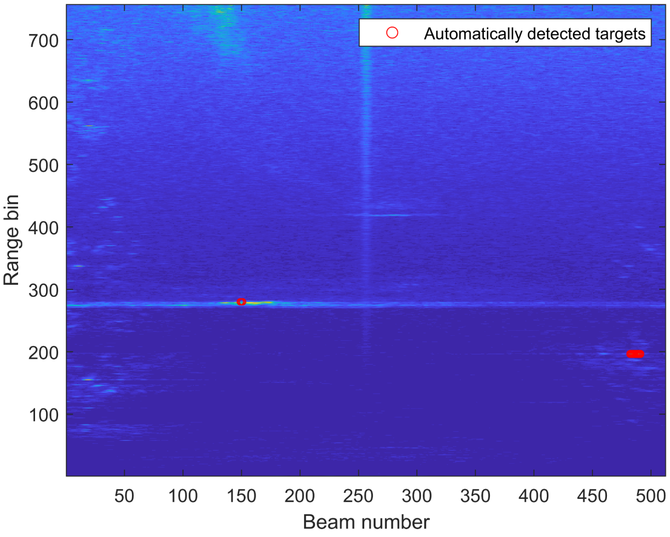


Figure 2.3.7b: Same image frame with pixels containing candidate targets circled in red.

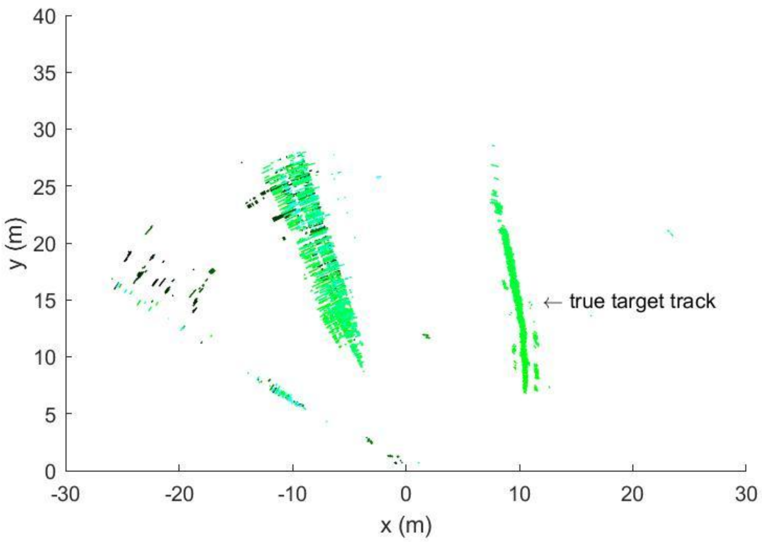


Figure 2.3.8: Time superposition of automatically detected candidate targets in Cartesian coordinates

2.3.3.2 Temporal filtering and sequence(s) building

Temporal filtering is performed on frames containing candidate targets. Filtering thresholds include a minimum number of contiguous frames containing candidate targets, and the maximum number of contiguous empty frames (frames where no candidate target has been detected). From this, temporal sequences of contiguous frames containing candidate targets are built (Figure 2.3.9).

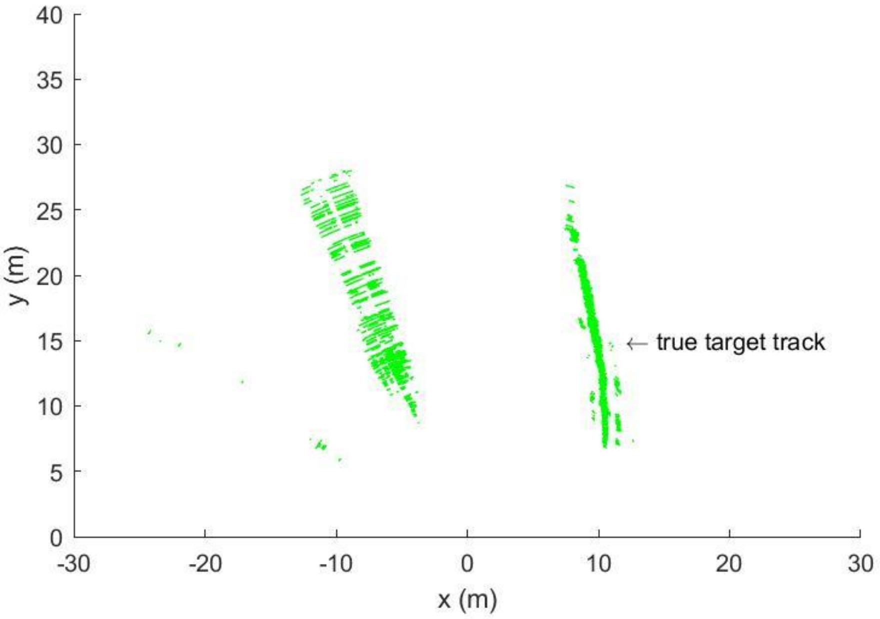


Figure 2.3.9: Time superposition temporal sequence of automatically detected candidate targets in Cartesian coordinates post temporal filtering

2.3.3.3 Target clustering and track reconstruction using Kalman Filter

For each frame within a temporal sequence, pixel coordinates containing detected targets are grouped into clusters. Cluster centroids are used as an input to a Kalman filter that reconstructs the track based on an estimate of the target's next location as seen in Figure 2.3.10.

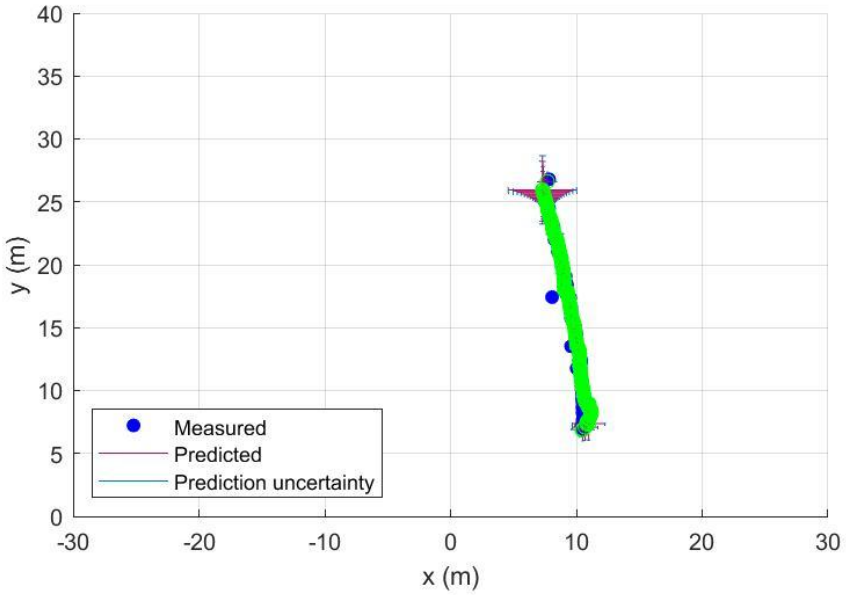


Figure 2.3.10: Track reconstruction using Kalman filter showing the automatically detected target coordinates parts of the target track along with the associated uncertainty

2.3.3.4 Exported results

For each dataset the following results are exported and automatically reported in a spreadsheet:

- Target frame number: Frame number that contains detected target(s) which can be corresponded to time, flow speed, etc.
- Target centroid: x,y coordinates (m) of the centre of the detected target(s) for each corresponding frame number.

2.3.3.5 Evaluation

The algorithm method has been evaluated against manual annotation of full length datasets.

As the detection and tracking approach (i.e., spatio-temporal, to yield target tracks) is different to the machine-learning approach (which searches manually identified frames for the target), hence, the evaluation criteria are also different.

Whilst automated detection returns the target frame number and the associated target(s) centre coordinates at an interval (period) of every frame, manual annotation of the dataset provides the target frame number and delineates the target(s) by defining a bounding box with upper and lower coordinates (m), at an interval (period) of 20 frames. A common base for temporal and spatial comparison was required to evaluate results.

Figure 2.3.11 shows the process for the temporal aspect of the comparison. (A) is the region of manually annotated target frames (black dotted lines) plus a buffer region of 40 frames on either side (~ two manual annotation intervals, or ~2s). Hence, if the manual annotations are considered a full and complete record of the target, then automated detections within the region (B) would be considered false temporal detections.

Figure 2.3.12 shows the process for the spatial aspect of the comparison. Again, (A) is the region of manually annotated target frames (black dotted lines) with a buffer region of 3 metres on either side. Thus automated detections within the region (B) would be considered false spatial detections.

To assess the algorithm performance for target detection, the following metrics referred as “True positive”, “False positive” and “Non-detected” are defined below :

- True positive: Automatically detected target is within spatio-temporal range of manually annotated target (automatic detection in both time AND space regions (A) in Figures 2.3.11 and 2.2.12).
- False positive: Automatically detected target is outside spatio-temporal range of manually annotated target (automatic detection in area (B) in time AND/OR space in Figures 2.3.11 and 2.2.12).
- Non-detected (false negative): No automatically detected target within spatio-temporal range of manually annotated target (No automatic detection in area (A) in both time AND space in Figures 2.2.10 and 2.2.11).

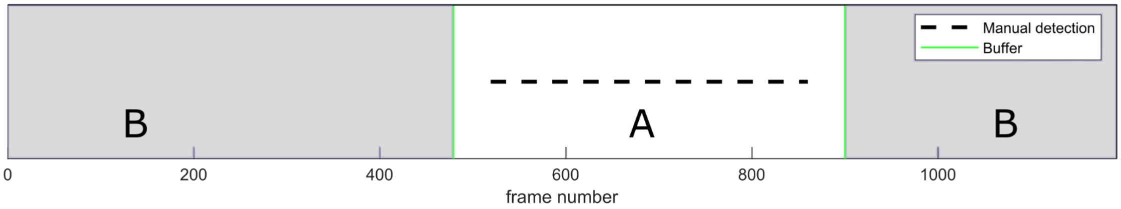


Figure 2.3.11: Process for temporal comparison of automated detections against manual annotations. (A) is the temporal period of manually annotated target frames (black dotted lines) plus a buffer region on either side. Automated detections at times (B) would be classified as false temporal detection.

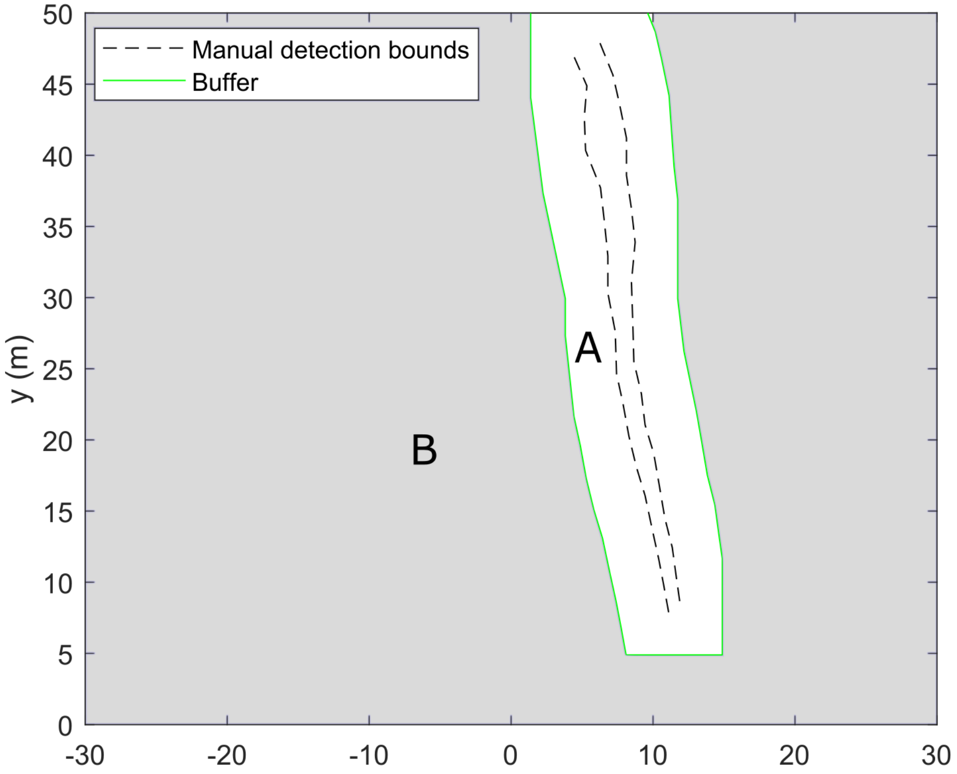
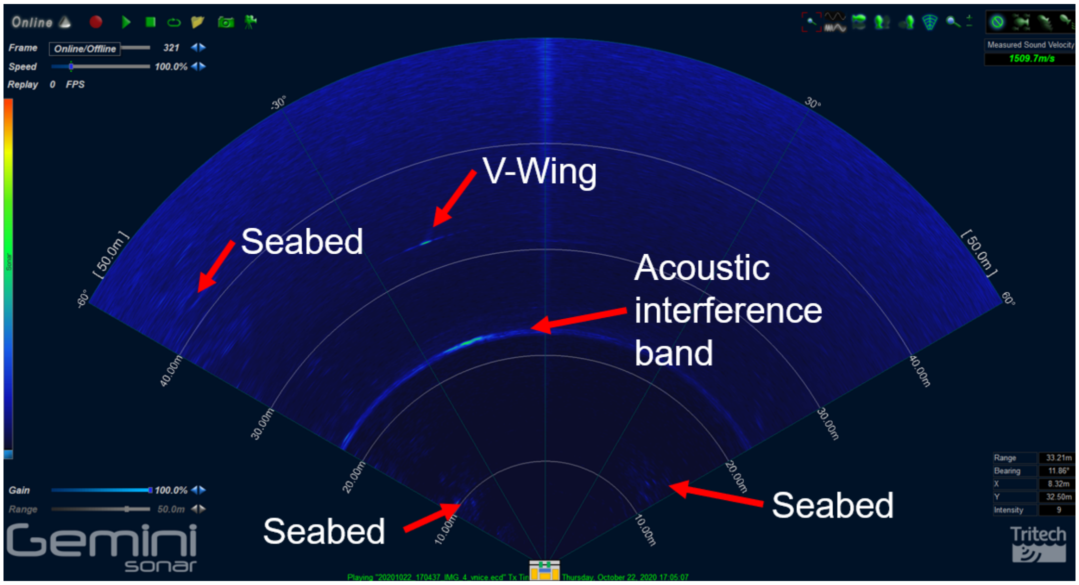


Figure 2.3.12 : Process for spatial comparison of automated detections against manual annotations. (A) is the region of manually annotated target frames, defined by the manual detection bounding box (black dotted lines) plus a buffer region on either side. Automated detections in region (B) would be classified as false spatial detection.

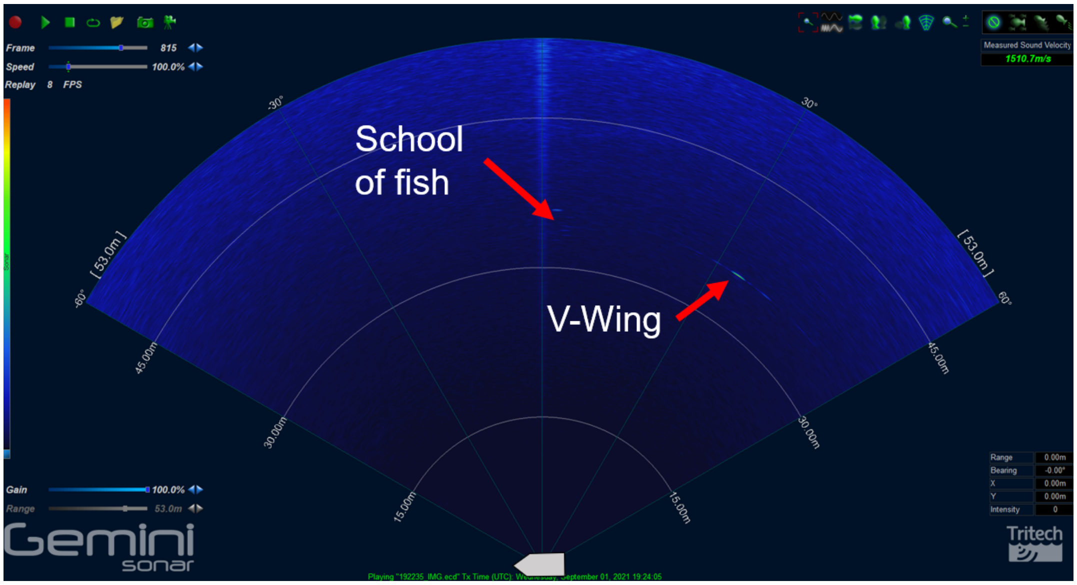
3 Results

The following subsections present the results of data analysis for the Grand Passage (bottom mount) and Minas Passage (vessel mount) datasets using automated detection with a) Gemini SeaTec software, b) optical-based deep learning methods, and c) spatial and temporal filtering. Example sonograms with returns from the V-Wing target are shown on Figures 3.1 and 3.2 for Grand Passage and Minas Passage, respectively.


Frame Online/Offline 321
Speed 100.0%
Replay 0 FPS
Gain 100.0%
Range 50.0m
Measured Sound Velocity 1509.7m/s

Range	33.21m
Bearing	11.86°
X	8.32m
Y	32.50m
Intensity	9

Figure 3.1: Example sonogram for Grand Passage with the V-Wing target


Frame 815
Speed 100.0%
Replay 8 FPS
Gain 100.0%
Range 53.0m
Measured Sound Velocity 1510.7m/s

Range	0.00m
Bearing	-0.00°
X	0.00m
Y	0.00m
Intensity	0

Figure 3.2: Example sonogram for Minas Passage with the V-Wing target

3.1 Gemini SeaTec

Target detection using the Gemini SeaTec software is used as a baseline for comparison to results from the a) optical-based deep learning, and b) spatial and temporal filtering methods discussed in sections 3.2. and 3.3.

As outlined in section 2.2.1, SeaTec uses an image processing blob detection method that identifies targets by differences in acoustic intensity, reflected by colours in the sonogram images. For each of the annotated frames in the Grand Passage and Minas Passage datasets we evaluated whether SeaTec detected the target of interest (true positive), missed the target of interest (false negative), and/or detected other returns (false positive). We performed our analyses with the default settings for sensitivity and target persistence. The following results should be interpreted with the understanding that improvements could be possible with effort applied to tuning the sensitivity and target persistence parameters.

3.1.1 Grand Passage

For the Grand Passage dataset we evaluated SeaTec's detection capabilities using the V-Wing target only. There were 468 total manual annotations of the V-Wing, of which 291 were detected resulting in a recall of 0.62, meaning that 62% of the V-Wing targets were detected. SeaTec detected an additional 2,228 targets in the annotated frames resulting in a precision of 0.12. That is, of all targets detected, 12% were the V-Wing. Many of the additional detections (false positives) were associated with a) acoustic interference from the Blueview sonar that was run concurrently with the Gemini and b) bottom returns. A summary of results is provided in Table 3.1.1.

Grand Passage	
	V-Wing
Total	468
True Positive	291
False Negative	177
False Positive	2228
Recall	0.62
Precision	0.12

Table 3.1.1: Grand Passage results for SeaTec detection of V-Wing target

3.1.2 Minas Passage

For the Minas Passage data set we evaluated SeaTec detection capabilities using the V-Wing, rock, and lead weight targets. A summary of results is provided in Table 3.1.2 showing a significant decrease in recall with decreasing target size. For the V-Wing there is increased performance in comparison to the Grand Passage dataset. This may be the result of a "cleaner" data set without acoustic interference or bottom returns. Many of the additional detections (false positives) were associated with the vessel used for target drifts and/or the rope for suspending the target. As such, these are associated with limitations of experiment methodology, and results for naturally occurring targets of similar size to the V-Wing may be improved.

Minas Passage			
	V-Wing	**Rock**	**Lead**
Total	686	182	124
True Positive	543	39	13
False Negative	143	143	111
False Positive	1337	167	23
Recall	0.79	0.21	0.10
Precision	0.29	0.19	0.36

Table 3.1.2: Minas Passage results for SeaTec detection of V-Wing, rock, and lead weight targets

3.2 Target Detection using Optical-based Deep Learning methods

In this section, we compare results from data curated from Grand and Minas Passages using the RetinaNet algorithm described in Section 2.3.2.2. We tested the performance of our trained models on several different variations of training data. In all cases, models were trained using the "retinanet_R_101_FPN_3x" configuration provided by FAIR. For Grand and Minas Passage testing, we look at several qualitative examples of correct and incorrect detection. We quantitatively evaluate the RetinaNet performance at these sites using the evaluation criteria noted in Section 2.3.2.3. Finally, we test model robustness by quantifying how well the model performs when testing data is noticeably different from training data. We test this by varying site and target between the train and test sets, to infer how well a trained model could be utilized in a different site and/or with a different object of interest.

3.2.1 Grand Passage

Grand Passage data contained interference bands and surface reflections alongside real target detections (Fig 2.3.4). The presence of noise, however, did not greatly impact the model's ability to detect targets in this environment. Below are some qualitative examples of correct detections from Grand Passage.

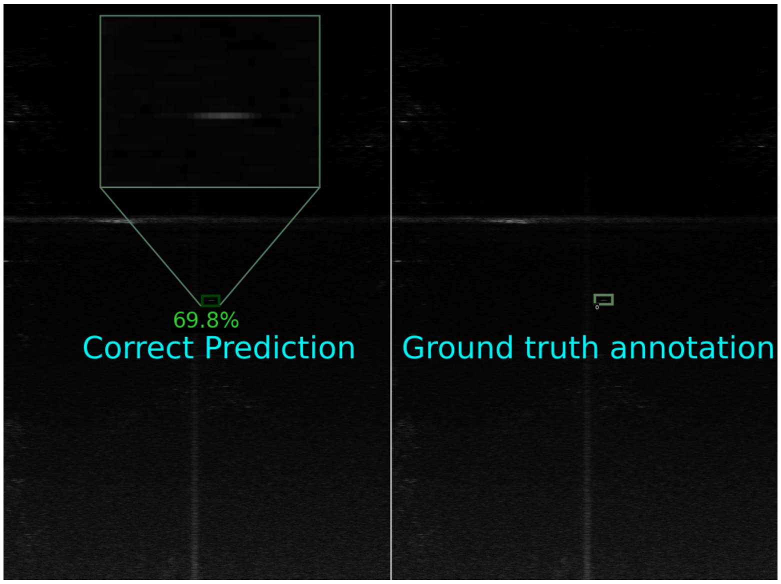


Fig 3.2.1. Grand Passage detection (left) and ground truth (right). Zoomed in picture of target shown in upper left frame.

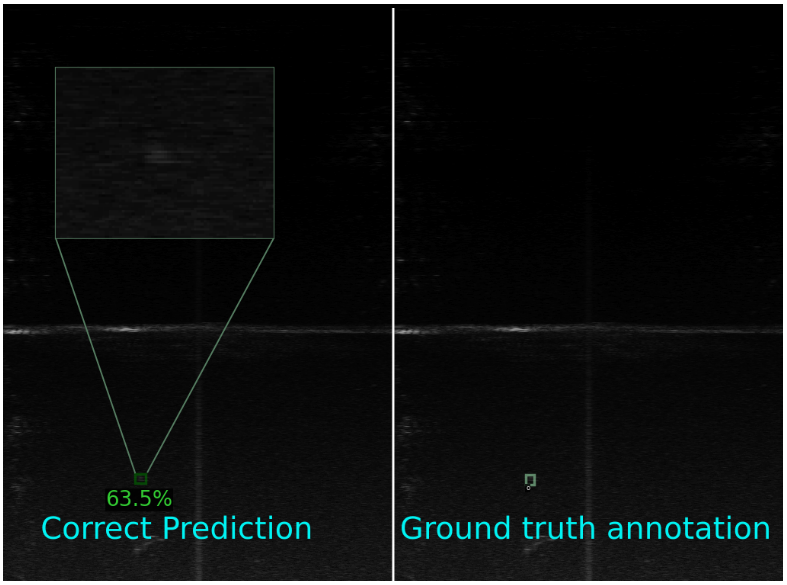


Fig 3.2.2. Grand Passage detection (left) and ground truth (right). Zoomed in picture of target shown in upper left frame.

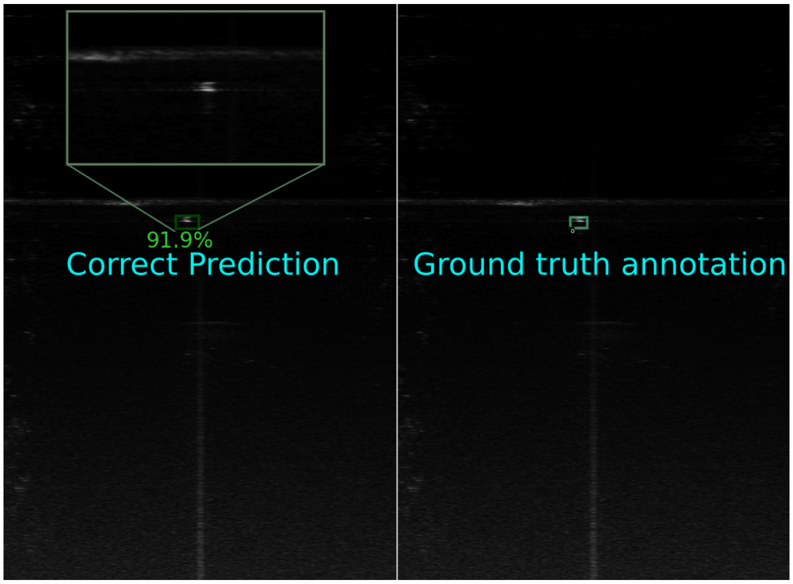


Fig 3.2.3. Grand Passage detection (left) and ground truth (right)

Several instances of bad detection results did occur, however. This included:

- Failing to detect targets (false negatives)
- Predicting targets that did not exist (false positives)
- Correctly predicting a target AND predicting targets which did not actually exist (true positives and false positives)

This first instance (missing targets completely) was typically the rarest of these events. This was partially by design because, as discussed prior, we consider missing true targets to be significantly more damaging than producing false positives. However, several cases did exist, particularly when considering individual target misses instead of whole-frame misses. For the latter two cases (the presence of false positives with and without true positive detections), this is likely due to setting our confidence threshold to levels lower than typical, resulting in low-confidence predictions occurring. In practice, this issue could be mitigated by 1) raising confidence threshold, at the cost of an increased false negative rate, or; 2) the inclusion of more (high quality) training data, to increase object detection confidence.

Examples of these failure cases are shown below, in addition to the failure case shown in Fig. 2.3.6:

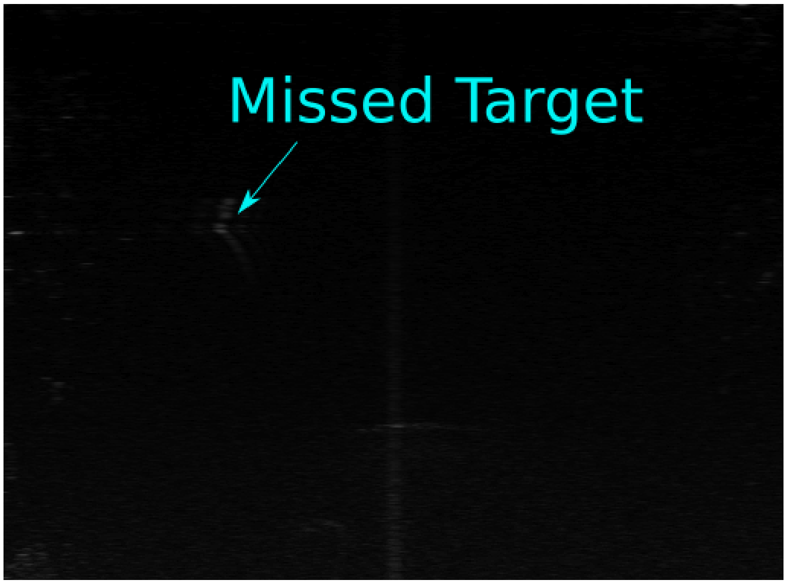


Fig 3.2.4. *Example failure case #1 at Grand Passage. Target (V-Wing) missed by detector. This was a rare case, but did occur at times.*

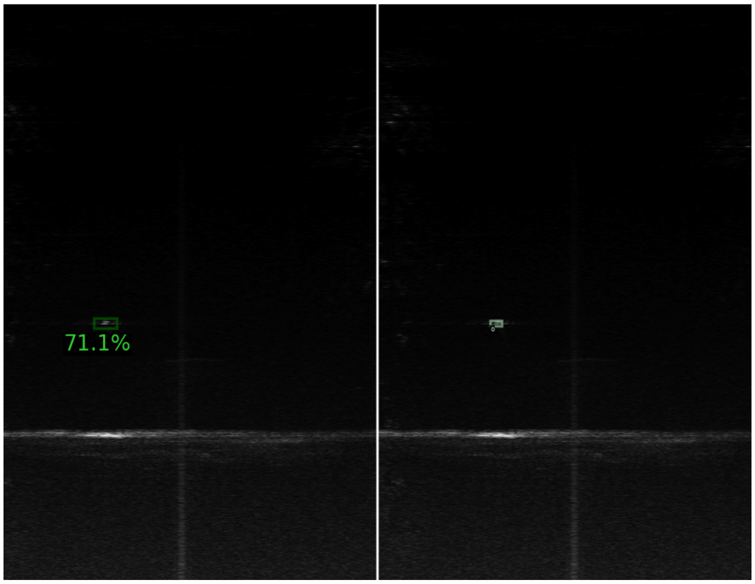


Fig 3.2.5. *Example failure case #2 at Grand Passage. Right: Ground truth target. Left: Predicted target. Predicted target's bounding box is too large compared to ground truth, resulting in poor IoU and therefore a false positive. A lower IoU threshold would result in a true positive result. This example shows how IoU can manipulate performance results, as most people would likely classify the prediction as correct in this case.*

Despite these mentioned failure cases, our quantitative analysis suggests that the utilization of RetinaNet allows for high frame-recall detection of target objects in the acoustic image.

IoU recall is lower, but remains encouragingly high. These Grand Passage results are shown in Tables 3.2.1 and 3.2.2 when trained/tested on all training data (e.g., without target filtering) and when only trained/tested on V-Wing data.

Results are presented for different confidence and IoU threshold levels. When trained/tested on all data (table 3.2.1), we see that this model exhibited high frame-recall with values above 98% in all cases (this indicates that 98+% of *frames* which contained a true positive target were detected by the model). IoU recall was lower at 80-84% (indicating that ~80% of *targets* were found), but still represents an encouraging figure. Precision measurements, on the other hand, are significantly worse than recall, at levels around 45-50%. This low precision was suspected to some degree, given the low confidence threshold utilized. In practice, precision can be raised by increasing the confidence threshold, at the cost of false negatives.

When limiting to just V-Wing data (table 3.2.2) (the most plentiful of the training data and the most pronounced in the acoustic image), we found similar results, but with an increased precision with decreased recall. This difference in result is likely due to removing many candidate 'fish' detections, which were not as pronounced as the V-Wing. Increased precision is therefore likely due to the model being better able to focus on the clearer, more obvious V-Wing targets. Decreased recall is a bit more curious, however. This may be due to lower levels of training data, impacting model convergence. Or, it could be due to random factors, where the model erroneously converged to an improper local minimum.

Confidence Threshold	**IoU Threshold**	**Precision**	**Frame Recall**	**IoU Recall**
0.3	0.3	0.491	**0.988**	**0.840**
0.4	0.4	0.467	0.987	0.8

Table 3.2.1: Grand Passage results for when not filtered for targets. Frame recall is shown as being greater than 98% in all test cases, with precision around 50%. IoU recall is noticeably lower, but still above 80%.

Confidence Threshold	**IoU Threshold**	**Precision**	**Frame Recall**	**IoU Recall**
0.3	0.3	0.616	**0.906**	**0.819**
0.4	0.4	0.592	0.902	0.787

Table 3.2.2: Grand Passage results when limited to V-Wing data. Frame recall is shown as being greater than 90% in all test cases, with precision around 60%. IoU recall is noticeably lower, but still above 75%.

The above results are encouraging; detecting ~90+% of frames with targets using automated processes is likely to be a boon to review processes. However, we caution that the large-scale adoption of this approach to more general datasets is unlikely to yield results as favourable as these, as we will discuss in Section 4.2.2.

3.2.2 Minas Passage

Evaluation of Minas Passage data followed the same procedure as Grand Passage, where a RetinaNet model is trained from labelled data sets. An example detection is shown in Fig 3.2.6; as can be seen, data from this site tended to be cleaner (i.e., with lower quantities of noise).

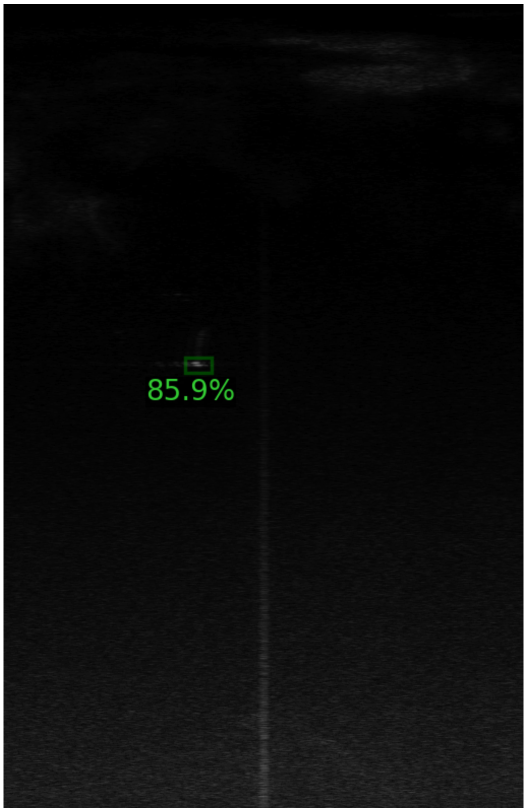


Figure 3.2.6: Detection at Minas Passage using a trained RetinaNet model. Imagery consists of considerably less interference than Grand Passage data e.g., Fig 3.2.2.

Following the same evaluation procedure as Grand Passage, we test the performance of this model at different IoU and confidence levels. Qualitatively, we find similar success and failure conditions as Grand Passage data, and omit extensive imagery for brevity. In general, we see a slight increase in model performance when testing at Minas Passage. This could be due to the less-noisy Minas Passage imagery, or due to the stochastic nature of training the model.

Quantitatively, we report results from this site in Tables *3.2.3* and *3.2.4*, again with and without ‘filtering’ the object of consideration.

Confidence Threshold	IoU Threshold	Precision	Frame Recall	IoU Recall
0.3	0.3	**0.642**	**0.823**	**0.652**
0.4	0.4	0.577	0.807	0.587

Table 3.2.3: Minas Passage results when trained on all data (not filtered for object). Frame (~80%) and IoU (58-65%) recall are both lower than Grand Passage. Precision (64%), however, is higher. The differences here compared to Grand Passage may be due to variations in data quality, or simply random noise with training the model.

Confidence Threshold	IoU Threshold	Precision	Frame Recall	IoU Recall
0.3	0.3	**0.761**	**0.931**	**0.807**
0.4	0.4	0.682	0.923	0.722

Table 3.2.4: Minas Passage results when limited to V-Wing data. Frame recall is shown as being greater than 92% in all test cases, with precision above 68%. Recall drops to 0.722-0.807 if IoU recall is utilized. For Minas Passage, limiting training and testing to V-Wing data led to a significant improvement for all the tested metrics.

These results, like that of Grand Passage, are encouraging. The tested models are routinely finding more than 80% of frames with targets when trained on appropriate quantities and qualities of data. Further, IoU recall similarly sees a noticeable decline compared to frame recall, but remains at a relatively high value of 80% in the best tested scenario.

Interestingly, and in contrast to Grand Passage data, we found higher recall when limiting to just V wing data (Table 3.2.4). This is probably due to how much more pronounced the V wing is compared to the other targets (rocks and lead) with respect to noise and other reflections. For example, when training was limited to only non V-Wing data (Table 3.2.5), performance dropped significantly across the board, likely due to the less explicit nature of this data set.

Confidence Threshold	IoU Threshold	Precision	Frame Recall	IoU Recall
0.3	0.3	**0.556**	**0.556**	**0.395**
0.4	0.4	0.481	0.52	0.342

Table 3.2.5: Limiting training data to non V-Wing data (i.e., rocks and lead)

While performance does vary across different situations and test cases, the high-level takeaway from Grand and Minas Passage testing is that the trained models are - for the most part - able to effectively identify targets in acoustic imagery when low confidence levels are utilized.

3.2.3 Model Robustness

While the results presented earlier for Grand and Minas Passage are promising, they do not explore how well a model will perform if the training and testing data are significantly different. While the presented analysis did appropriately split data into testing and training partitions, both of these data are of the same general “type”: the same location, looking at the same target(s). Here, we purposely break this relationship and test model performance when 1) the *object* of interest is varied, while keeping the site the same, and; 2) the *site* is changed, while keeping the object of interest the same. While it is well understood that model efficacy will decrease when inference-time data is further removed from training data appearance, the *degree* to which this degradation will occur in the presented data is not well understood. Therefore, these two cases – robustness to target variation and robustness to site variation – are explored in this section.

3.2.3.1 Robustness to target variation

First, we explored how well a model could detect an object when the object type has been varied. For this test, we trained on V-Wing data from Minas Passage, and tested on non V-Wing data from the same site. The results are shown in Table 3.2.6, with a frame recall of 40% and an IoU recall of 31.5%. This is noticeably lower than models trained on the same target (Tables 3.2.2 and 3.2.5), and suggest difficulty when targets are varied. This is not particularly shocking, as detectors are not meant for inter-class detection. For example, an optical detector trained to detect cats would fail if tasked with detecting dogs. It’s likely we are seeing a similar case here. However, in the acoustic case, similarities between targets are much higher than e.g., the optical appearance of cats and dogs, so recall is not nearly as degraded as an analogous optical case where the algorithm would likely exhibit complete failure. Still, this result suggests that these targets are not similar *enough* for a detector to operate with high performance. Training with more representative data will typically be required for target detection. In section 4.2.2, we explore how adding *in situ* data to a model can improve model performance over a base model when no representative data is present.

Confidence Threshold	IoU Threshold	Precision	Frame Recall	IoU Recall
0.3	0.3	0.571	0.40	0.315

Table 3.2.6: Performance when trained on V-Wing data and tested on non-V-wing data. Train set: Minas Passage V-Wing. Test: Minas Passage non V-Wing

When the test is flipped - such that we train on non V-Wing data and test on V-Wing data, the result is relatively similar, as shown in Table 3.2.7.

Confidence Threshold	IoU Threshold	Precision	Frame Recall	IoU Recall
0.3	0.3	0.6444	0.414	0.349

Table 3.2.7: Train set: Minas Passage non V-Wing. Test: Minas Passage V-Wing

3.2.3.2 Robustness to site variation

Our second test considered inter-site comparison, where a detector was trained for one site, and tested on another. For this test, we limited training and testing data to V-Wing data, to limit the variations in data outside of site differences, to avoid conflating different potential sources of error.

First, we trained on Grand Passage data and tested on Minas Passage data, as shown in Table 3.2.8. Frame recall remained relatively high - greater than 95% - when trained on this site and one tested on Minas Passage. This suggests that a substantial majority of targets will be found using these detector types, even when data available for training is not site-specific. Logically, this makes sense: an optical detector which was trained to detect cats from images inside an apartment's living room is unlikely to fail if asked to detect cats in, say, the kitchen.

Confidence Threshold	IoU Threshold	Precision	Frame Recall	IoU Recall
0.3	0.3	0.701	0.983	0.779

Table 3.2.8: Train on Grand Passage data, test on Minas Passage data (V wing Only)

However, when we flip this test - train on Minas Passage and test on Grand Passage - we find significantly worse results, as Table 3.2.9 shows. We believe the reason for the stark differences between these two tables is due to the higher noise levels present at Grand Passage compared to Minas Passage. That is, when training on the relatively clean Minas Passage data, the model did not recognize the difference between noise and object as effectively as the Grand Passage model. Therefore, when testing occurred on the more noisy Grand Passage data set, it failed to predict targets to a reliable degree.

Confidence Threshold	IoU Threshold	Precision	Frame Recall	IoU Recall
0.3	0.3	0.427	0.508	0.3404

Table 3.2.9: Train on Minas Passage data, test on Grand Passage Data (V wing Only)

The contrast in these two results suggests that acoustic models may exhibit robustness to site. However, there is a limit to this performance, and significantly different locations may see substantially worse performance.

3.3 Detection and tracking through spatial and temporal filtering

In this section, results from the comparison between automated detection and tracking through spatial and temporal filtering against manual benchmark annotation are presented. Comparison has been performed on the datasets collected at Grand Passage in Section 3.3.1 and Minas Passage in Section 3.3.2.

3.3.1 Grand Passage

This dataset contains 31 manually annotated files, each containing one single track of the V-Wing target. Figures 3.3.1 and 3.3.2 show an example of an automatically detected target track marked as true positive. The detected track is in the manual temporal range (Figure 3.3.1) and spatial range (Figure 3.3.2), hence counted as a true positive.

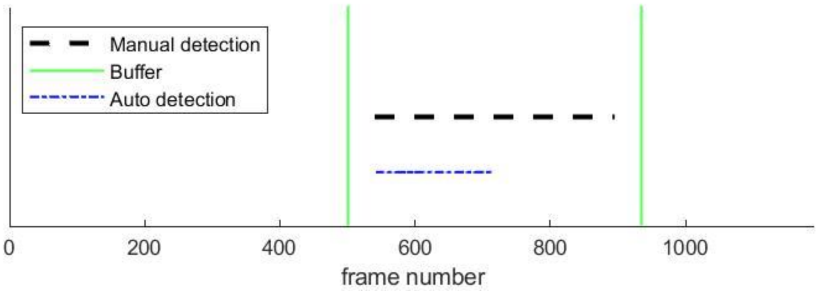


Figure 3.3.1 : Temporal superposition of manually and automatically detected frames containing targets, with the manually annotated frames in dotted black lines, temporal buffer in green and automatically detected frames in blue.

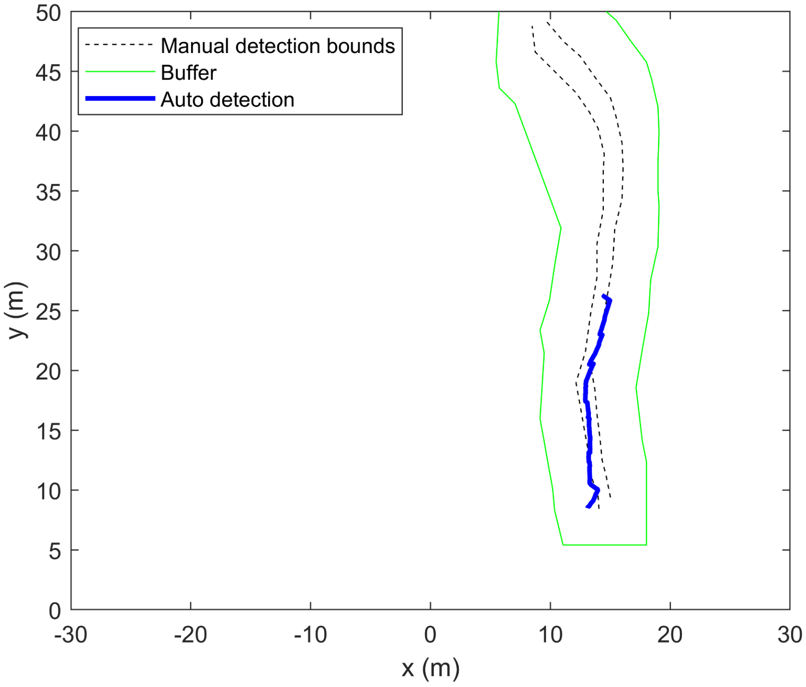


Figure 3.3.2 : Spatial superposition of manually and automatically detected targets with the manual detection bounds in dotted black lines, spatial buffer in green and automatically detected target track centres in blue.

Table 3.3.1 summarizes the results of the algorithm against manual annotations. On this dataset, the automatic detection contains 30 true positives and 2 false positives with 1 non-detected target. This corresponds to a Recall of 97% with a Precision of 94%. Both high precision and recall show that the algorithm can enable automated detection of large datasets.

Total number of targets	True positive	False Positive	Non detected	Precision	Recall
31	30	2	1	0.938	0.968

Table 3.3.1: Grand Passage summary of automated target detection on the entire dataset

3.3.2 Minas Passage

This dataset contains 45 annotated files, that includes 28 V-Wing targets, 10 rock targets and 7 lead targets. As described in section 2.1.2, the V-Wing is significantly larger than the rock and lead target (52-cm V-Wing, 12-cm diameter rock and 3.8-cm lead). Hence for the files containing rock and lead targets, thresholds of the algorithm have been adjusted to increase detection sensitivity. This enables more true detections (higher recall) but at the expense of an increased number of false positives (lower precision), given the overall objectives of regulatory-acceptable automated detection of targets, and that automated detection even with this level of precision is many orders of magnitude less labour intensive than manual review. In addition to this, reflections from the rope are often visible and detected as a viable target by the algorithm. This is illustrated in Figure 3.3.3, which shows the time superposition of potential targets detected by the step 1 of the algorithm, with colour gradient changing as function of time. The true target track is visible in (1) at approx. 19-m range along with the reflections from the rope in (2) between 10-m and 15-m range. The algorithm was not specifically adjusted to exclude the rope target, as this artificial target is not representative of eventual deployment of the sonar or algorithm in a tidal stream site for biological targets - hence detection and tracking sensitivity was preserved.

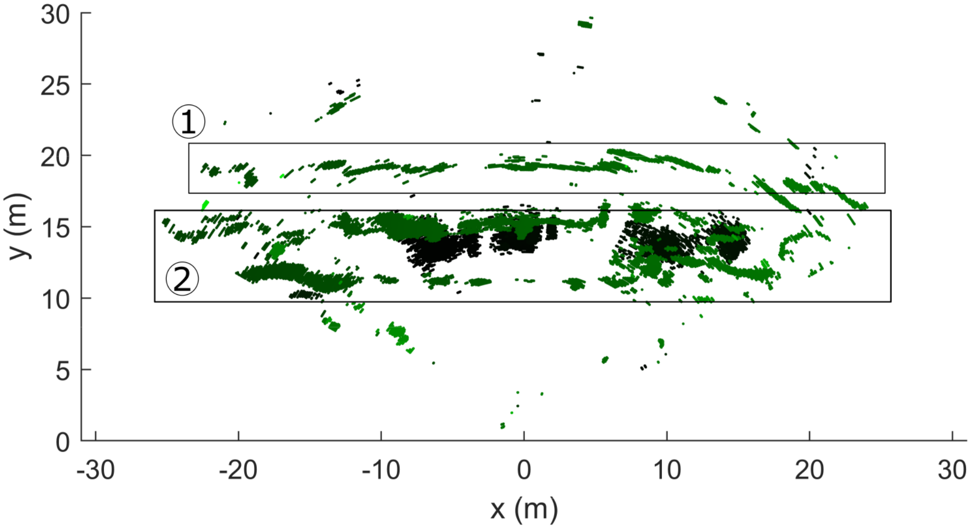


Figure 3.3.3: Time superposition of potential targets detected by the step 1 of the algorithm, showing the true target track (1) and the reflections from the rope (2) .

Tables 1, 2, and 3 summarize the detection results and performance of the algorithm for the datasets containing V-Wing, rock and lead targets respectively. Recall varies from 82% to 100%. The high number of false positives leading to a low precision is due to the reflections from the rope or occasional vessel wake being misinterpreted as targets by the algorithm (these were within the manual annotation temporal range but outside the spatial range). These tables show the capability of the algorithm to detect targets independently of their types (size, shape, etc.) in a complex environment.

Total number of V-Wing targets	True positive	False Positive	Non detected	Precision	Recall
28	23	49	5	0.319	0.821

Table 3.3.2: Minas Passage result summary of automated target detection of V-Wing targets. Low precision values were due to detection of the rope suspending targets, i.e., valid detection of water-column returns, unlikely to be present in a final dataset and a classification problem rather than detection/tracking.

Total number of rock targets	True positive	False Positive	Non detected	Precision	Recall
10	9	35	1	0.205	0.900

Table 3.3.3: Minas Passage result summary of automated target detection of rock targets. Low precision values were due to detection of the rope suspending targets, i.e., valid detection of water-column returns, unlikely to be present in a final dataset and a classification problem rather than detection/tracking.

Total number of lead targets	True positive	False Positive	Non detected	Precision	Recall
7	7	49	0	0.125	1.00

Table 3.3.4: Minas Passage result summary of automated target detection of lead targets. Low precision values were due to detection of the rope suspending targets, i.e., valid detection of water-column returns, unlikely to be present in a final dataset and a classification problem rather than detection/tracking.

4 Case Study

4.1 Pathway Monitoring Platform

As part of NTZ's Pathway Project a cabled instrument frame (the Pathway Monitoring Platform) was deployed at the FORCE site by Halagonia Tidal Energy Limited (HTEL). The deployment location is shown in Figure 4.1 and the Pathway Monitoring Platform is shown in Figure 4.2. More information on the deployment and data collection can be found in the Pathway Monitoring Platform Project Final Report (Thomas, 2022).

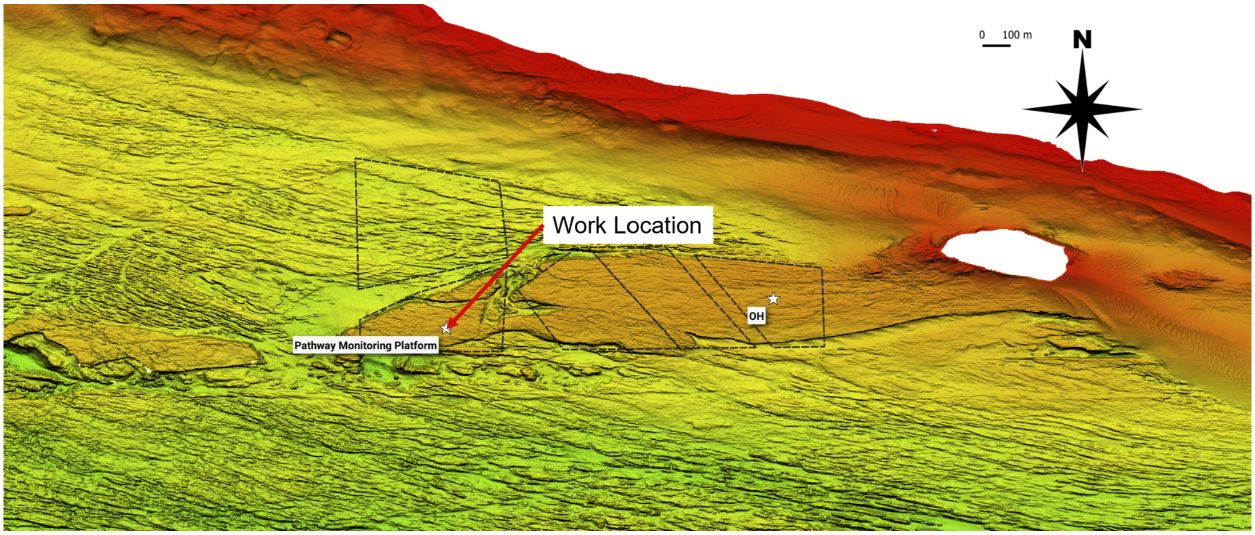


Figure 4.1: Pathway Monitoring Platform deployment location

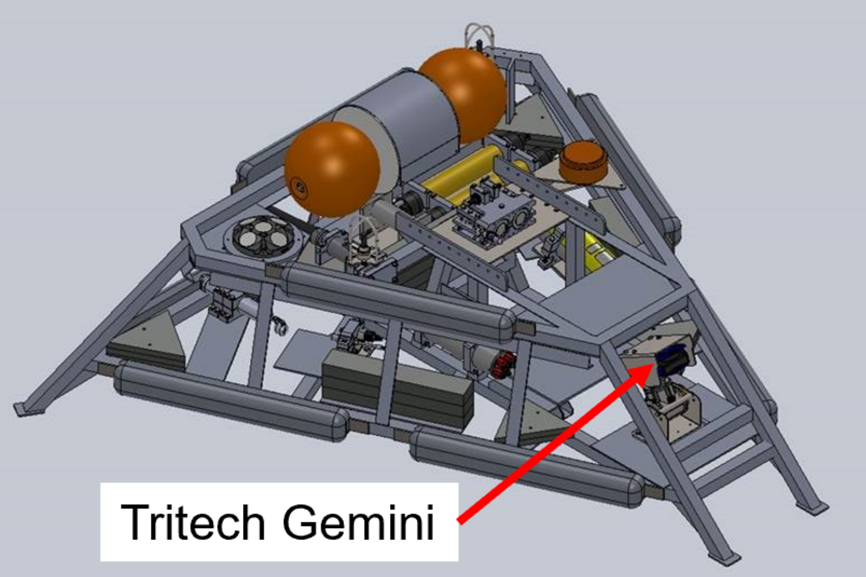


Figure 4.2: Pathway Monitoring Platform with Tritech Gemini mounted in the pan and tilt module on the bottom right.

HTEL collected continuous data with the Tritech Gemini from February 7, 2022 at 16:30 UTC to February 8, 2022 at 10:49 UTC without any additional active acoustic instruments pinging. These data were supplied to our project team for initial analysis as a case study to evaluate the performance of the automated analysis.

4.2 Data Analysis

4.2.1 Gemini SeaTec

The Gemini SeaTec software developed by Tritech was used for manual data analysis, including annotating target detection and tracks for training and testing automated analysis.

Consistent with methodology for other data sets (as described in Section 2.2.1) annotations included the file name, time, ping number, a bounding box for the target in Cartesian coordinates (lower X, lower Y, upper X, and upper Y), target description, and general notes. Approximately 200 annotations were logged in a Google Sheet used to facilitate sharing amongst the project team, and easy export as .csv format for programmatic training and evaluation of approaches for automated analysis.

Analysis was comprehensive for the first hour of data supplied by HTEL, with all targets identified manually being annotated for at least three points on their track. The targets were generally small and often had a low level of contrast in acoustic intensity relative to the background noise in the sonogram. Optical validation is not practical in the Minas Passage due to visibility generally limited to less than 3 m and often less than 1 m, though the targets appeared to be typical of individual fish. Manual detection was greatly reliant on the presence of target tracks, that is, the visual persistence of the targets in the sonograms between consecutive time steps.

The Gemini SeaTec target detection feature was evaluated by SOAR. The software returned many false positives and was often not able to detect the targets of interest while using the highest level of sensitivity. This is consistent with the result shown in Section 3.2.1, for which the use of SeaTec led to low recall scores for the rock and lead weight targets. However, there is potential to improve results by excluding regions with high levels of noise from the analysis. An example sonogram is shown in Figure 4.3 with a false negative (target not detected) and greater than 500 false positives. Most of the false positives would be removed if the maximum range was set to 40 m, however the false negative result would remain.

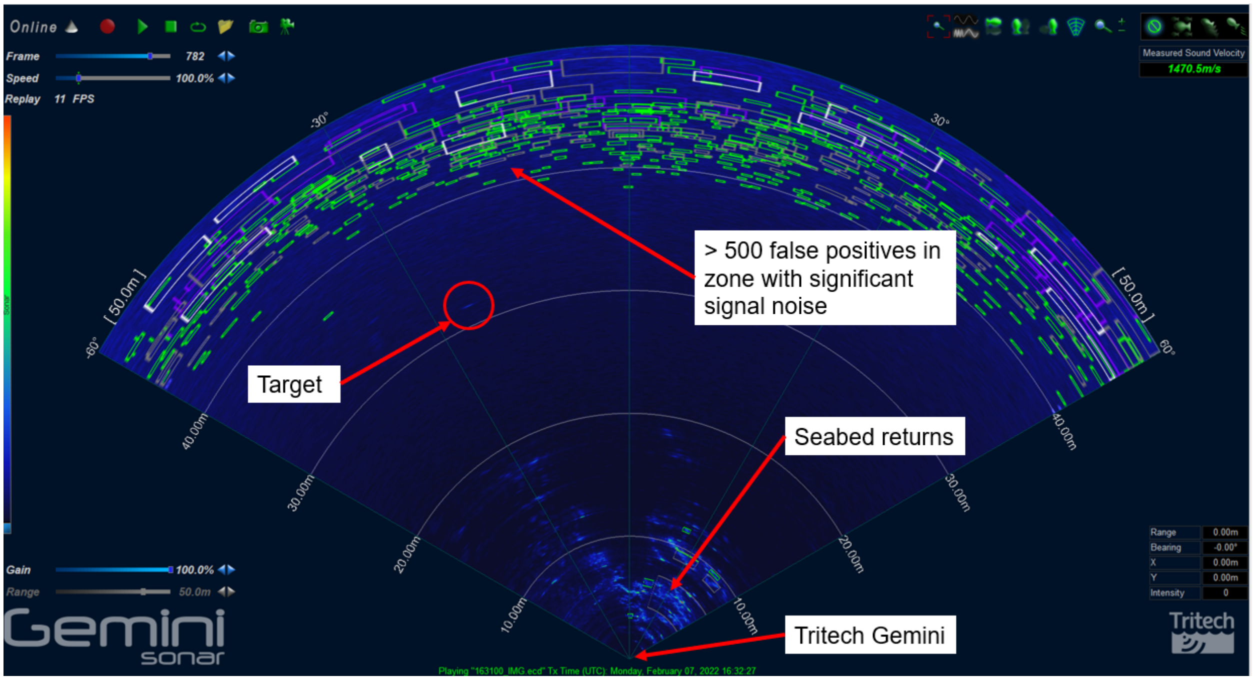


Figure 4.2: Pathway Monitoring Platform example sonogram showing faint target, seabed returns, and significant false positives associated with signal noise in the 40 to 50 m range zone.

4.2.2 Target Detection using Optical-based Deep Learning methods

For a first pass, the target detection using optical deep learning approach utilized models trained for the Grand and Minas Passage. We found extremely poor results, with recalls below 10%. This result is somewhat contrary to our findings in Section 3.2.3, where we found significant model robustness when utilizing a Grand Passage model on Minas Passage data. However, as discussed in 3.2.3, model performance will likely be very poor when the site is significantly different (e.g., different noise properties) and/or with different targets of interest. Therefore, the most likely reasons for this poor result may be due to the factor: 1) searching for both a different target within a different environment, 2) the new environment is drastically different in appearance from the train environment, 3) the target of interest was noticeably different at inference time than train time (live fish vs e.g., towed V-Wing). This poor result was likely a combination of all these factors. Additionally, the targets in the data set are very small, faint, and difficult to distinguish from other noise and background. This small size makes transferring models from Grand Passage and minas passage challenging, while also complicating performance of site-specific models. Examples of this site are shown in Figs 4.2.2 and 4.2.3.

Despite this difficult environment, training this model using correct-site data led to relatively strong performance. Results for this site are shown in Table 4.2.1.

Confidence Threshold	IoU Threshold	Precision	Frame Recall	IoU Recall
0.3	0.3	0.639	0.821	0.657

Table 4.2.1: Performance of model when trained on HTEL dataset without any transfer-learning from Minas or grand passage weights

As is shown, this method is capable of detecting ~82% of frames which contain biological targets in the environment at these low IOU thresholds. While not as the performance seen at Grand and Minas Passages, this result is still encouraging for the practical uses of machine learning algorithms in real world scenarios. Capturing and annotating more data for model training and reducing the acoustic interference at the site will likely improve the performance of these real world scenarios.

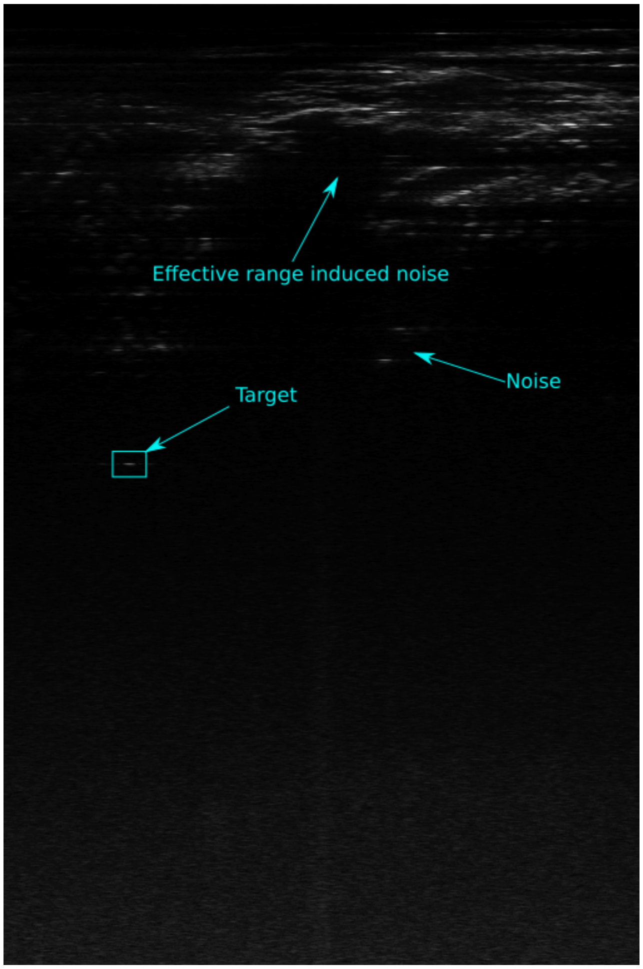


Fig 4.2.2: Example HTEL data image. A target, in the middle left, is shown alongside noise and seafloor reflections. To the human eye, the target is difficult to distinguish from noise. Note that this is one of the clearer examples of targets in this dataset.

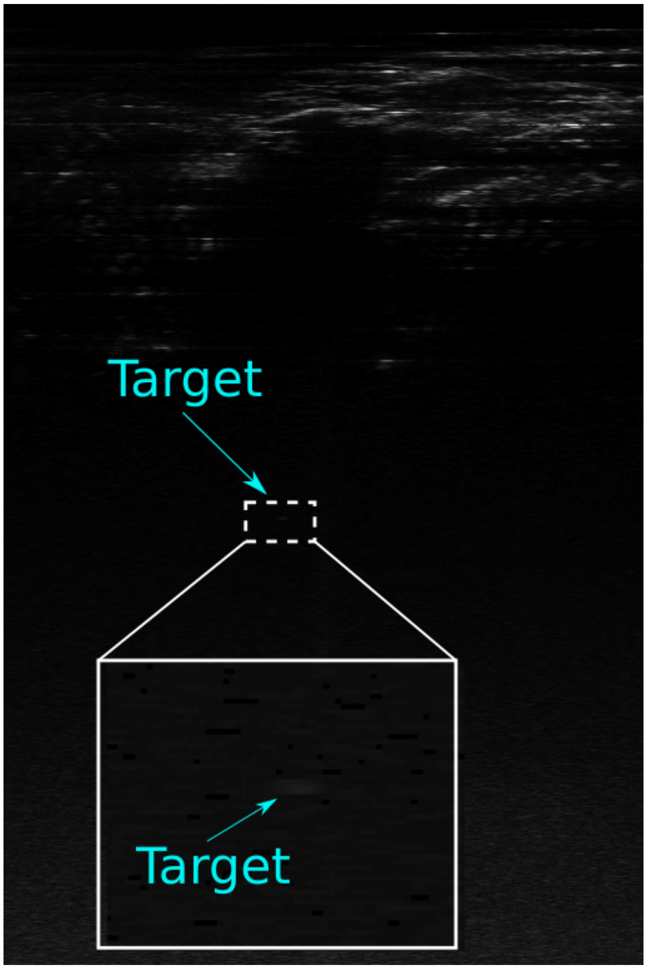


Fig 4.2.3: Example HTEL data image. The target - in the centre of the image - is very faint and difficult to distinguish from noise and the image background. This represented a typical case for targets in this dataset.

However, the results from table 4.2.1 were based on model weights initialized from the 'standard' RetinaNet weights provided by FAIR. That is, the weights were initialized on optical data weights, before being refined for the specific task. If we utilize data from Grand and Minas Passage - all available data, without filtering for targets - we find improved performance (Table 4.2.2), with frame recall at 88.9%.

DP Data Amount	Confidence Threshold	IoU Threshold	Precision	Frame Recall	IoU Recall
150	0.3	0.3	0.649	0.889	0.686

Table 4.2.2: Performance of model when trained on HTEL dataset with weights initialized from weights trained from Grand and Minas Passage data. Frame recall is shown at 88.9%.

This increase in performance utilizing more generic acoustic imagery is encouraging, as it suggests that non *in situ* data may be helpful in training a more site specific model. While the increase in frame recall is not extremely significant - 82% vs 89% - it does represent that model performance can likely be improved by leveraging previously trained models, which may reduce the time and data required to adapt a model to a specific site.

An important question, then, is what quantity of *in situ* data is required to improve performance to desired levels for a specific environmental task. We aimed to briefly study that in Table 4.2.3, where the quantity of DP Energy utilized at train time is increased in small quantities. We do find that, outside of IoU recall for 120 frames, recall improves as the quantity of data increases in this training scheme. This follows logically, as adding more representative data should improve model performance. The drop in IoU recall at 120 frames could be due to random factors with model training, or because the 40 frames added between the 80 and 120 cases were poorly labelled cases, or cases which were not particularly representative.

A few other interesting trends can be found from Table 4.2.3. First, precision does not seem to increase over training. This is interesting, and suggests the model may have converged to near maximum precision using this approach. Second, recall does not appear to 'level off', suggesting that this model would be further improved with more data. Finally, performance does not appear to be better than the 'standard' model (i.e., Table 4.2.2 with weights initialized from FAIR weights and trained on DP energy data) until 120 *in situ* images are added. This likely means earlier data was overfit to the training data, a logical conclusion based on the small number of training data utilized.

DP Data Amount	**Confidence Threshold**	**IoU Threshold**	**Precision**	**Frame Recall**	**IoU Recall**
20	0.3	0.3	0.645	0.667	0.589
60	0.3	0.3	0.667	0.786	0.629
80	0.3	0.3	0.605	0.793	0.657
120	0.3	0.3	0.6111	0.846	0.629
150	0.3	0.3	0.649	0.889	0.686

Table 4.2.3: Adding small quantities of in situ HTEL data and training with weights initialized from Grand and Minas Passages. In general, recall increases, while precision seems to remain about the same.

A challenge encountered in this dataset arose from dealing with the effective range of the sonar. The effective range is a function of multiple factors including the frequency of the sonar (as attenuation increases with frequency) and the amount and type of scatters in the field of view. For this particular dataset we found the effective range to be approximately 30 m, resulting in significant noise in the top of the image (Fig 4.2.2).

While out of scope for the presented study, limiting targets to only consider targets outside of this noisy zone would likely improve model performance. We believe this study should be a subject of future work.

4.2.3 Detection and tracking through spatial and temporal filtering

The detection algorithm was run blindly on the dataset and compared thereafter with manual annotations performed on the first 35 files.

Figure 4.2.4 shows an example of time superposition of candidate targets automatically detected at step 1 of the algorithm (prior to noise removal). Background and strong seabed returns have been filtered out, and visible target tracks can be seen (numbered from 1 to 5).

Figure 4.2.5 and 4.2.6 show the superposition of manually detected tracks and automatically detected target tracks that met all the algorithm criteria in spatial (a) and temporal space (b). Note that the spatial buffer (in green) has been increased to 5 m on both sides of the manual detection area to take into account the slight shift in target coordinates between the manual annotation and target detection.

It can be seen that tracks 1 and 3 have been both manually annotated and automatically detected. Tracks 2 and 4 were not manually annotated in spite of meeting the automatic detection criteria. Track 5 was not manually annotated nor saved as a definite target by the algorithm.

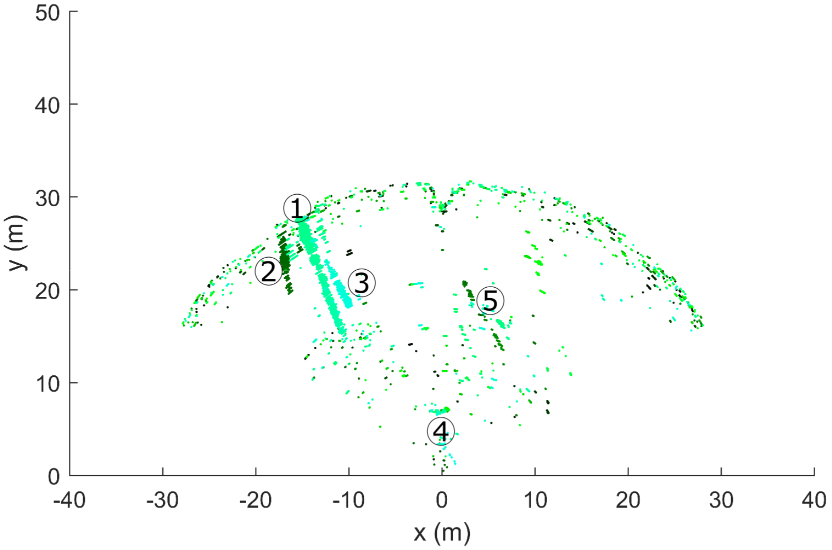


Figure 4.2.4: Time superposition of automatically detected candidate targets after the first step of the algorithm (prior to noise removal), showing visible target tracks numbered from 1 to 5.

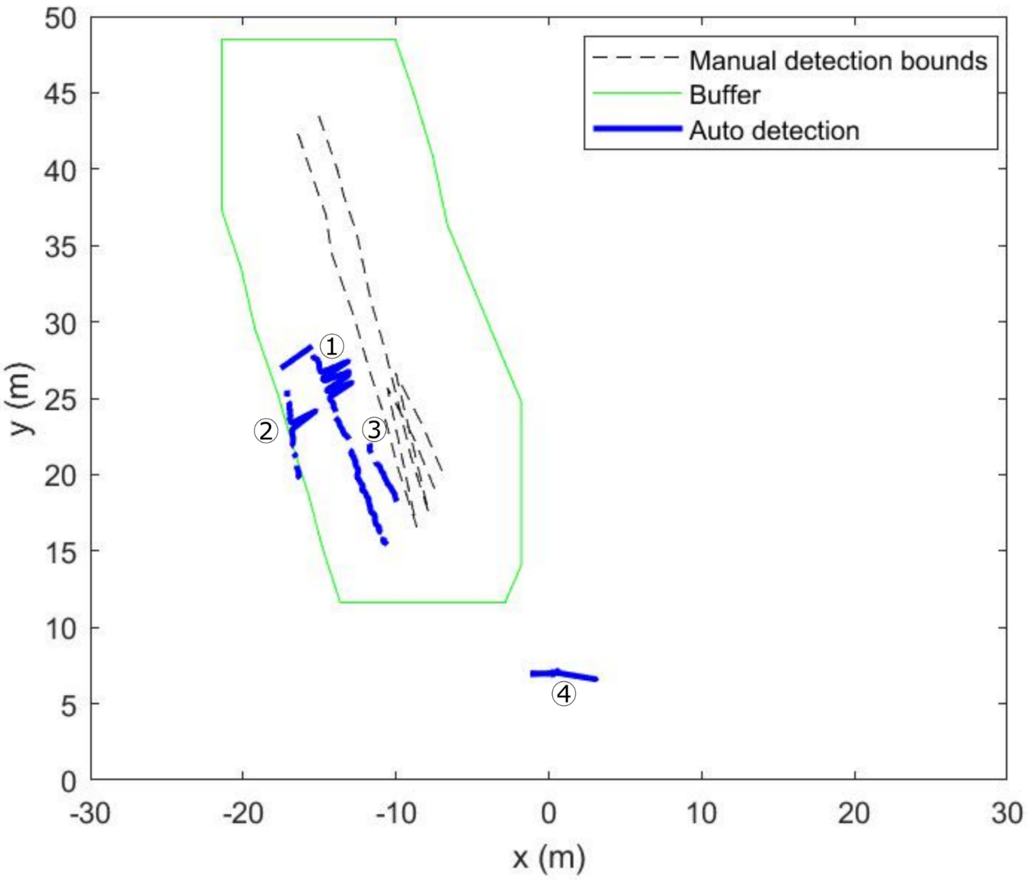


Figure 4.2.5 : Spatial superposition of manually and automatically detected targets with the manual detection bounds in botted black lines, spatial buffer in green and automatically detected target track centres in blue and numbered from 1 to 4.

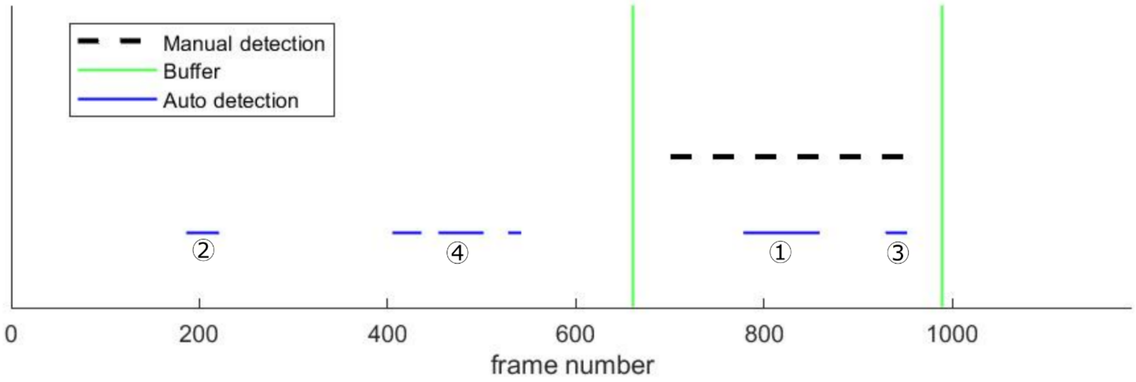


Figure 4.2.6 : Temporal superposition of manually and automatically detected frames containing targets with the manually annotated frames in dotted black lines, temporal buffer in green and automatically detected frames in blue and numbered from 1 to 4. .

In this case study, tracks detected within the spatio-temporal range of manual annotations are marked as true positives as in Sections 2.2.3.5, while tracks detected outside the spatio-temporal range of the manual annotations are marked as flagged for further manual investigation.

Table 4.2.4 presents a summary of the comparison results between the automated detection and manual annotations. Of the 24 manually detected targets, 18 were detected by the algorithm. This corresponds to 75% of true detection of single or pairs of fish, whose returns were annotated from strong to very weak. In addition to this 33 additional target tracks have been flagged for further manual investigation which were not annotated by the reviewer, potentially demonstrating increased sensitivity of detection.

True positives	Tracks flagged for further manual investigation	Non detected
18	33	6

Table 4.2.4*: DP platform data result summary of automated target detection compared against manual annotations*

5 Conclusions and Recommendations

Three methods for automated analysis of data from the Tritech Gemini multibeam imaging sonar have been evaluated utilizing artificial targets deployed in Grand Passage and Minas Passage. The targets were annotated manually, then used to train and evaluate: a) image processing "blob detection" applied using the Gemini SeaTec software, b) optical-based deep learning methods over each frame annotated with a target and evaluate c) detection and tracking using spatial and temporal filtering over the entire dataset. Hence, for the UHI tracking algorithm, additional targets detected and tracked which were not annotated by the reviewer were classified as 'false positives' resulting in a lower 'Precision' in Table 5.1(*); this demonstrates detection in an entire dataset, rather than just detection of a target within frames annotated to contain a target. For further details see Section 2.3.3.5 and Figures 2.3.11 and 2.3.12.

The metrics of precision and recall were used to evaluate the accuracy. Precision indicates the portion of all predicted targets which were true targets, i.e.:

$$P = \frac{tp}{tp + fp}, \quad \text{Eqn. 1}$$

For precision P, true positive tp, and false positive fp. Recall, which is also known as the true positive rate, evaluates what portion of targets in a database where found by the algorithm:

$$R = \frac{tp}{tp + fn}, \quad \text{Eqn. 2}$$

for recall R and false negative fn. True positives tp and false positives fp are calculated only within frames annotated as containing targets when using SeaTec and optical-based deep learning (OBDL) methods, and across the entire dataset when using UHI detection and tracking algorithm/filter. A summary of results is provided in Table 5.1, and each method is discussed individually within the following sections.

Metric	Method	Minas Passage			Grand Passage
		V-Wing	Rock	Lead	V-Wing
Precision	**SeaTec**	0.29	0.19	0.36	0.12
	OBDL	**0.76**	**0.56**	-	0.62
	UHI Filter	0.32*	0.20*	0.12*	**0.94**
Recall	**SeaTec**	0.79	0.21	0.10	0.62
	OBDL	**0.93**	0.56	-	0.91
	UHI Filter	0.82	**0.90**	**1.00**	**0.97**

Table 5.1*: Summary of results. *For the UHI algorithm, additional targets detected and tracked which were not annotated by the reviewer were classified as 'false positives' resulting in a lower 'Precision'; this demonstrates detection in an entire dataset, rather than just detection of a target within frames annotated to contain a target.*

5.1 Gemini SeaTec

Throughout our work, including during previous stages (Trowse et al., 2020 and 2021), the Gemini SeaTec software has proven to be reliable for instrument setup and data collection with a user-friendly interface. Though the instrument's software writes proprietary Gemini .ecd files, the raw data can be accessed programmatically for conversion and potential compression into alternate formats for storage and analysis. Although lossless conversion was used in this project, any conversion or compression should carefully consider losses of data quality which can significantly affect data analysis.

Consistent with the previous experiments, the effective range of the Gemini sonar was found to be 30 m to 40 m depending on size of target and environmental conditions (bubbles, sediment, zooplankton, and other acoustic scatterers). Beyond this effective range targets are not easily visible amongst the background noise.

The recall scores achieved using SeaTec's target detection and tracking module with the V-Wing were greater than 60% and 70% for the Grand Passage and Minas Passage data sets, respectively. However, the module also returned many false positives, indicating a possible need for manual review of the data (low precision). Recall was reduced significantly with smaller targets including the rock (21%) and lead weight (10%). For biological targets of interest, SeaTec detection could be expected to perform reasonably well for detection of large targets with temporal persistence and high contrast relative to the acoustic intensity of the background (e.g., marine mammals, sharks, and schools of fish). SeaTec was not found to be useful for detection and tracking of the small individual potential fish targets in the case study data set provided by HTEL. This was not an unexpected result, given the reliance of blob detection-type methods on adequate target contrast and persistence.

5.2 Target Detection using Optical-based Deep Learning

The presented work explored the efficacy of one algorithm - RetinaNet - to detect candidate targets in multibeam sonar imagery. This algorithm, which we trained as a *binary* detector to avoid challenging inter-class differentiation, was developed for and is primarily utilized on optical data. However, as the work shows, the algorithm is capable at detecting target objects in multibeam sonar data at low confidence. The caveat is that precision is lower than ideal, but recall - particularly *frame* recall, where target localization is not considered - remains encouragingly high. This algorithm may offer high enough recall to support autonomous target identification.

However, as with all deep learning approaches, the quality and quantity of input training data is of extreme importance. This work reinforced that naïve utilization of models trained outside of test data constraints offers noticeably worse recall and precision. This is especially true when training data targets are noticeably different in acoustic appearance than testing data targets. However, we found that 1) this decrease in model performance may be less pronounced than optical data, likely due to extreme similarity in the acoustic signature of different targets, and 2) these models exhibit some robustness to site variation unless there is an extreme change in noise characteristics between sites.

While the presented work is encouraging, the authors stress that the adaptation of this (or similar) model to live data requires further study and integration development. Namely, the topics of model robustness, appropriate quality and quantity of data, data ‘cleaning’, model augmentation with *in situ* data, and inter-algorithm comparison should be considered (including novel methods which are specifically aimed at acoustic data). Additionally, while frame recall is high, the same cannot be said for precision. Lower precision levels are due to deliberately allowing low confidence predictions to occur to minimize the likelihood of missing targets.

Missing targets completely was rare. This was partially by design because, as discussed prior, we consider missing true targets to be significantly more damaging than producing false positives. However, several cases did exist, particularly when considering individual target misses instead of whole-frame misses. For the latter two cases (the presence of false positives with and without true positive detections), this is likely due to setting our confidence threshold to levels lower than typical, resulting in low-confidence predictions occurring. In practice, this issue could be mitigated by 1) raising confidence threshold, at the cost of an increased false negative rate, or; 2) the inclusion of more (high quality) training data, to increase object detection confidence.

Future work should explore if the same level of recall can be achieved but with higher confidence and precision.

5.3 Detection and tracking through spatial and temporal filtering

Performance of the “Detection and tracking through spatial and temporal filtering” algorithm has been assessed on non-biological targets in Section 3.3 and applied to a tidal stream dataset containing biological targets in Section 4.2.3. The algorithm was shown to be successful at automated detection and tracking of targets from a raw dataset, i.e., application to a continuous data stream, providing identification of frames containing targets and tracking movement of targets over time, a complementary approach to the Optical-based Deep Learning algorithm for target detection within annotated frames.

Results from the datasets containing non-biological targets showed a recall of >96% for Grand Passage data with a precision of 94%, and a recall from 80 to 100% for Minas Passage data with a low precision due to the rope suspending non-biological targets being identified as a target by the algorithm – given the algorithm was designed to detect targets in the water column distinct from the background, detection of the rope is to be expected and unlikely to be in an eventual tidal stream turbine monitoring application. If e.g., mooring ropes were present in an eventual application, these can be masked to improve precision. Importantly, the algorithms were also shown to be effective at detecting and tracking smaller targets i.e., rock and lead (Table 5.1).

The algorithms were tuned to provide a balance between recall (offering regulator certainty) and precision (providing several orders of magnitude lower volumes of data for review, facilitating continuous long-term monitoring). Trials of the algorithm on the DP platform dataset demonstrated encouraging results with 75% true detection of single or pairs of fish. Further detections were marked for additional manual verification as these presented the characteristics of a potential fish track. Importantly, the algorithm provided greater sensitivity than manual annotations, identifying several fish tracks that were not annotated by the reviewer, while preserving robustness identifying tracks rather than spurious detections of noise.

For improved performance, the algorithm thresholds can be tuned as a function of the dataset characteristics (i.e. presence of interference bands, or returns from the seabed). Adaptive filtering preserves sensitivity of detections across flow speeds and across the swath range of the sonar. Sensitivity of the algorithm can also be increased towards further detection of smaller targets. For a fixed instrument deployment (e.g., sonar on a turbine or seabed platform), these thresholds only need to be adjusted once. The algorithm can then run continuously on large datasets, without the need for human input, providing automated outputs of: target detection time, detected frames numbers, and target coordinates, i.e., a track from which regulator metrics such as encounter rate, evasion, movement and behaviour can be derived. This is particularly useful on large datasets for continuous monitoring, such as collected by the DP platform, where manual annotations are not feasible for continuous monitoring.

Automated detection and tracking was also demonstrated from boat-based deployments, where environmental noise (wakes, seabed returns, etc.) requires greater adaptive temporal and spatial filtering to discriminate targets from background.

Whilst the algorithm was successful at detecting tracks of all target sizes, biological and non-biological, annotated and not-annotated, in different environments and deployment configurations, it should be noted that in the far-range (>30 m) of the sonar, the start or end of the tracks were often missing. This is due to the target intensity decreasing into an increasingly cluttered background. This is a natural limitation of the diverging swath of multibeam echosounders (sonar) and dependent on deployment configuration due to returns from the seabed / sea surface. Hence, the algorithm provided detection and tracking – the original requirement – yet for the greatest measures of animal behaviour, track extension would be beneficial. This can be optimised by swath orientation, and future work will aim to improve detection in the far-range, to extend detected tracks using adaptive sensitivity for track extension with the Kalman filter. This could also be investigated using the Optical-based Deep Learning method trained on the previously detected part of the track.

To summarize, use of the detection and tracking algorithm through spatial and temporal filtering enables a rapid detection screening that returns both frames containing targets and coordinates of detected targets within those frames. This enables an automated pre-screening of the data, pointing the user directly towards the frames flagged as containing a potential target, drastically reducing operator input from manual review (precision) while preserving a cautious approach to target detection and tracking in light of regulatory priorities (preserves recall). This tool becomes particularly useful on large datasets or continuous monitoring, where manual review of data rapidly becomes unfeasible.

5.4 Combined approach

Finally, future work should explore the possibility of combining the deep learning-based detection algorithm with filtering techniques such as the "Detection and tracking through spatial and temporal filtering" approach presented in this work. We see two potential ways these approaches can be combined. In the first, the spatial and temporal filtering algorithms can be used to generate annotations for model training of a deep-learning algorithm. In the second, a deep-learning approach could be used to augment a spatial and temporal filtering algorithm for detection and tracking; for example, where the current spatial and temporal filtering algorithm is able to identify tracks, but the extreme start/end is limited by signal-to-noise at swath periphery, increased and adaptive sensitivity to target detection in a focused area could be feasible with a deep-learning approach utilizing the existing frame-based approach, i.e., to identify the bounding box locations of candidate track extensions.

5.5 Transferability of techniques

With widespread use of multibeam imaging sonars to monitor fish and marine mammal presence and behaviour in tidal stream sites, and to provide robust continuous nearfield monitoring around turbines irrespective of visibility or illumination, there is wide applicability to the techniques demonstrated here. The tracking and detection algorithms are tunable for different target characteristics to provide robust detection of different species, and the adaptive filtering approach is robust across hydrodynamic conditions of different sites, including the two locations and deployment configurations demonstrated here, and building on other tracking at the MeyGen and EMEC sites in Scotland (Williamson et al., 2021; Couto et al., 2022). The algorithms have also been shown to be transferable between multibeam imaging sonars of different resolutions, building on and extending previous work with a lower-power lower-resolution multibeam imaging sonar (Williamson et al., 2021). Real time processing capability, and parameterisation of raw datastreams (TB / day) into robust detections and target tracks have been shown, together with demonstrating automated detection sensitivity beyond that of a manual observer. Additional validation data through future demonstration, including across different sites and conditions, will support further transferability of the techniques.

6 References

(Couto, 2022) A. Couto, B.J. Williamson, T. Cornulier, P.G. Fernandes, S. Fraser, J.D. Chapman, I.M. Davies, B.E. Scott (2022). Tidal streams, fish, and seabirds: Understanding the linkages between mobile predators, prey, and hydrodynamics. Ecosphere. http://doi.org/10.1002/ecs2.4080

(Huang 2017) Huang, Jonathan, et al. "Speed/accuracy trade-offs for modern convolutional object detectors." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

(Lin 2017a) Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *Proceedings of the IEEE international conference on computer vision, 2017.*

(Lin 2017b) Lin, Tsung-Yi, et al (2017b). "Feature pyramid networks for object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

(Neupane 2020) Neupane, Dhiraj, and Jongwon Seok. "A review on deep learning-based approaches for automatic sonar target recognition." *Electronics* 9.11 (2020): 1972.

(Rosebrock 2016a) Adrian Rosebrock. A photo of a stop sign with two bounding boxes: ground-truth and prediction. *Wikimedia Commons.* 2016. https://commons.wikimedia.org/wiki/File:Intersection_over_Union_-_object_detection_bounding_boxes.jpg

(Rosebrock 2016b) Adrian Rosebrock . A visual equation for Intersection over Union (Jaccard Index). *Wikimedia Commons.* 2016 https://commons.wikimedia.org/wiki/File:Intersection_over_Union_-_visual_equation.png

(Singh 2021) Deepak Singh and Matias Valdenegro-Toro. The marine debris dataset for forward-looking sonar semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 3741–3749, 2021.

(Thomas 2022) S. Thomas 2022. Pathway Monitoring Platform Project Final Report. Final Report to the Net Zero Atlantic. Halagonia Tidal Energy Limited. Halifax, NS, Canada.

(Trowse 2020) G. Trowse, T. Guest, G. Feiel, R. Cheel, and Hay, A.E. 2020. Field Assessment of Multibeam Sonar Performance in Surface Mount Deployments. Final Report to the Offshore Energy Research Association of Nova Scotia. SOAR – Sustainable Oceans Applied Research, Technical Report. Freeport, NS, Canada.

(Trowse 2021) G. Trowse, T. Guest, G. Feiel, R. Cheel, and Hay, A.E. 2021. Field Assessment of Multibeam Sonar Performance in Bottom Mount Deployments. Final Report to the Offshore

Energy Research Association of Nova Scotia. SOAR – Sustainable Oceans Applied Research, Technical Report. Freeport, NS, Canada.

(Valdenegro-Toro 2019) Valdenegro-Toro, Matias. "Deep neural networks for marine debris detection in sonar images." *arXiv preprint arXiv:1905.05241* (2019).

(Williamson 2021) B.J. Williamson, P. Blondel, L.D Williamson, B.E. Scott (2021). Application of a multibeam echosounder to document changes in animal movement and behaviour around a tidal turbine structure. ICES Journal of Marine Science.
http://doi.org/10.1093/icesjms/fsab017\