EXECUTIVE SUMMARY

Sonar-Based Deep Learning for Underwater UXO Remediation – Phase II

David Williams
NATO STO Centre for Maritime Research and Experimentation (CMRE)

May 2021
EXECUTIVE SUMMARY
Project: MR18-1444

TABLE OF CONTENTS

1.0 INTRODUCTION ................................................................................................................. 1
2.0 OBJECTIVES ........................................................................................................................ 1
3.0 TECHNICAL APPROACH ................................................................................................... 1
4.0 RESULTS AND DISCUSSION ............................................................................................ 3
  4.1 DATA NORMALIZATION ......................................................................................... 3
  4.2 DETECTION ................................................................................................................ 4
  4.3 CLASSIFICATION ...................................................................................................... 7
  4.4 LIMITED-SCOPE CLASSIFICATION ...................................................................... 8
5.0 IMPLICATIONS FOR FUTURE RESEARCH AND BENEFITS ..................................... 10
6.0 REFERENCES .................................................................................................................... 10

LIST OF FIGURES

Figure ES-1. An Example SVSS Volumetric Scene Image Displayed as a Trio of MIPs, When the Data is (a) Raw or (b) Normalized......................... 3
Figure ES-2. Performance of the Detection Algorithm for Each SVSS Data Collection for (a) Proud Man-made Targets and (b) Buried Man-made Targets, along with the Distribution of Visual Human Assessment Ratings. ....... 5
Figure ES-3. (a)-(d) SVSS Alarm Cubes (Each Displayed As a Trio of MIPs) of Four Targets Extracted from Figure ES-1, and (e)-(h) Photographs of the Objects During Installation (pre-Burial). ....................................................... 6
Figure ES-4. Example Alarms Generated by the Detection Algorithm when Applied to a MuST Scene................................................................. 6
Figure ES-5. Overall Performance of the Detection Algorithm for Each SVSS Data Collection ........................................................................ 7
Figure ES-6. Classification Performance on the SVSS Test Data Set in Terms of ROC-like Curves for (a) Only Proud Objects and (b) Only Buried Objects........ 8
Figure ES-7. For Different Input Data Representations, Classification Performance for Discriminating Air-filled and Water-filled Objects Using 155 mm Munitions as Training Data........................................................................... 9
<table>
<thead>
<tr>
<th>ACRONYMS AND ABBREVIATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>APL-UW</td>
</tr>
<tr>
<td>ATR</td>
</tr>
<tr>
<td>AUC</td>
</tr>
<tr>
<td>BRAC</td>
</tr>
<tr>
<td>CNN</td>
</tr>
<tr>
<td>DoD</td>
</tr>
<tr>
<td>FUDS</td>
</tr>
<tr>
<td>MIP</td>
</tr>
<tr>
<td>MuST</td>
</tr>
<tr>
<td>PSU-ARL</td>
</tr>
<tr>
<td>ROC</td>
</tr>
<tr>
<td>SERDP</td>
</tr>
<tr>
<td>SVSS</td>
</tr>
<tr>
<td>TIERSWAT</td>
</tr>
<tr>
<td>UXO</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENT

The author thanks the following:

- the Strategic Environmental Research and Development Program (SERDP) for funding this project.
- Daniel Brown at the Penn State University Applied Research Laboratory (PSU-ARL) for providing the Sediment Volume Search Sonar (SVSS) data.
- Timothy Marston at the Applied Physics Laboratory - University of Washington (APL-UW) for providing the Multi-Sensor Towbody (MuST) data.
- Kevin Williams at APL-UW for providing the TIERSWAT data.
1.0 INTRODUCTION

An unfortunate legacy of former military activities at sites designated for base realignment and closure (BRAC) and at formerly used defense sites (FUDS) is the contamination of aquatic environments with military munitions. In the United States, more than 400 underwater sites, spanning an area in excess of 10 million acres, potentially contain such munitions. The presence of these munitions is a serious threat to both humans and the environment, so remediation is necessary. But the return of these contaminated waters to public use is contingent upon the analysis and assessment of wide-area and detailed underwater surveys. Therefore, the Department of Defense (DoD) has an express need for the development of technologies that will enable the detection and classification, at high probability, of military munitions found at underwater sites.

2.0 OBJECTIVES

After the successful execution of SERDP Exploratory Development project MR18-1444 [1], this project was in response to SERDP Statement of Need “MR19-1444-F1: MR18- 1444 Follow-on,” in the Munitions Response Program Area. Specifically, the research fell under the topic of “Wide Area and Detailed Surveys,” by addressing the need for technologies that would enable the detection and classification, at high probability, of military munitions found at underwater sites.

The primary objective of this project was to develop novel UXO detection and classification algorithms specifically for volumetric sonar data from two experimental systems, the SVSS developed under SERDP project MR-2545 (PI: D. Brown) and the MuST from ESTCP project MR18-5004 (PI: K. Williams). An auxiliary objective was to explore the use of limited-scope experiments within a CNN framework in order to improve the explainability of the resulting deep-learning classifiers.

Developing the detection and classification algorithms was expected to fill a capability gap, and as a consequence, enable more efficient remediation efforts at contaminated sites.

3.0 TECHNICAL APPROACH

Because no automatic target recognition (ATR) algorithms previously existed for these two new systems, the methods developed here addressed a capability gap. The general-purpose detection algorithm that was created exploited the concept of integral images to flag suspicious regions in a given data volume in a fast, computationally efficient manner. The follow-on classification algorithm was based on deep-learning techniques, specifically deep convolutional neural networks (CNN) that were extended to function with three-dimensional (i.e., volumetric) input data cubes. The developed algorithms were assessed using large sets of SVSS data, and they were also applied to modest amounts of data from the MuST system. Preliminary results showed the promise of the approaches for detecting and classifying both proud and buried objects in measured volumetric sonar data.

This project made extensive use of data collected by the SVSS [2] and MuST [3, 4] systems to aid in the development and assessment of various ATR algorithms. The systems are complementary in the sense that each focuses on different water depth regimes. The SVSS system is designed to support UXO remediation in shallow water (1-5 m), while the MuST system is intended to operate in deeper water (6-40 m). Both systems employ low-frequency sonar to enable the production of volumetric SAS imagery, with this in turn facilitating the detection of proud and buried objects.
Because the two systems produce roughly comparable data, an effort was made to design flexible algorithms that could function with data from either system (as well as other similar systems developed in the future). Eschewing system-specific algorithms also makes the exploitation of data from multiple systems in a joint manner – e.g., for classifier training – more feasible.

The main algorithms developed for the SVSS and MuST data addressed normalization, detection, and classification. Additional experiments were conducted for limited-scope CNN classifiers using simulated acoustic-color sonar data. Here the project team provided extremely cursory summaries of these new algorithms.

Given a raw 3-d data cube of beamformed sonar returns, the normalization procedure developed for volumetric sonar data is as follows. First determine the plane of the interface in the image volume for which returns dominate. Then determine and remove the sub-volume of the data cube contaminated by multipath interference associated with the dominant interface. For a given cross-track position and a given distance from the dominant interface, normalize by the median value. Similarly, for a given along-track position and a given distance from the dominant interface, normalize by the median value. Finally, perform a logarithmic transformation.

The object detection task is a classic remote-sensing problem of locating a target signal amid a noisy background [5]. The proposed algorithm computes a local estimate of the background intensity around each voxel, which are the potential target signals. If the target-to-background ratio exceeds a set threshold, the voxel is flagged. And if a connected volume of such flagged voxels exceeds the minimum size of objects of interest, a discrete alarm is generated. The summed intensity over a rectangular volume needed for the target or background estimates was computed in an extremely efficient manner using integral images [6].

A CNN [7] is a sophisticated classification algorithm that customarily ingests an image as input and produces scalar outputs corresponding to the probabilities of belonging to each class under consideration (e.g., target and clutter). The classification approach was based on 3-d CNNs in which the input data is a 3-d data cube, rather than a 2-d image. The general architecture design largely follows the previous work [8], but extends it to three dimensions. The project team also chose to employ an ensemble of eight CNNs, each of which has a unique network architecture, so that complementary clues may be discovered and exploited by the CNNs. This in turn improves the overall robustness of the classification scheme. Despite the enormous size of the input data cubes, the use of extremely small networks allows the CNNs to be trained with modest computational resources, and vitally, avoids computer-memory constraint issues. A CNN-based classification approach is particularly apropos for this data modality because hand-crafting features is challenging, and CNNs effectively obviate this process.

The aforementioned CNNs were used to discriminate between two general classes of objects, each of which exhibits considerable intra-class diversity in terms of object size, shape, composition, and burial state. As a result, the features that a trained CNN will rely on to make a prediction will be a complex combination that is not easily disentangled. Instead, designing specially controlled experiments can provide a way to learn principled, explainable features that can be tied directly to the wave phenomena of the physics involved. The idea is to develop a CNN classifier in which the two classes are not UXO and non-UXO, but rather whether or not a specific object has a certain attribute. As a result, any clues that the CNN uncovers should be due to the single variable that differs.
Here the project team exploited this limited-scope experiment concept (using simulated acoustic-color data) to train CNNs to discriminate air-filled objects from water-filled objects.

### 4.0 RESULTS AND DISCUSSION

This section presents a selected set of results of the previously described algorithms.

#### 4.1 DATA NORMALIZATION

The challenge of visualizing a 3-d data cube often leads to the use of a 2-d maximum intensity projection (MIP), which collapses the imagery along one of its principal axes by retaining the highest intensity voxels along that axis [9]. A typical data cube from the SVSS system, before and after the proposed normalization algorithm, is shown in Figure ES-1 as a set of three 2-d MIPs in a common reference frame. In Figure ES-1(a), the dominant interface return obscures target signatures, whereas in Figure ES-1(b) it can be observed that the normalization procedure amplifies the signals of interest, including elastic target returns. Although not the principal objective, the normalization method also facilitates human interpretability of the data.

![Figure ES-1](image)

**Figure ES-1.** An Example SVSS Volumetric Scene Image Displayed as a Trio of MIPs, When the Data is (a) Raw or (b) Normalized.

*Algorithm detections are marked on the depth MIPs with red dots.*
4.2 DETECTION

For the results reported in this work, the SVSS system was used to collect data at three sites in the United States, two distinct locations in the Fosters Joseph Sayers Reservoir in Pennsylvania and one location at the Aberdeen Test Center in Maryland. At the Sayers sites, there existed an upper layer of approximately 8 cm of silt atop a clay base; site “A” had a 1.3 m water depth, while site “B” had a 3.0 m water depth. The Aberdeen location, site “C,” featured a sloping sediment of sand, resulting in a water depth that ranged from 1.0 m to 2.5 m. The shallow water meant multipath interference was not insignificant.

The sites were reservoirs that could be drained to facilitate target emplacement. So prior to data collection, various man-made objects were deployed, including aluminum cylinders, steel pipes, steel shot puts, concrete blocks, and an assortment of munitions with diameters ranging from 20 mm to 155 mm. Some objects were placed proud on the sediment, while others were buried to various depths (up to 60 cm) below the water-sediment interface. Data collections at Sayers took place in 2019 at the following times (nominally labeled “1” through “4,” respectively): June, August, early November, and late November. The Aberdeen collection (labeled “5”) occurred in March 2020.

The complex physics at play underwater and in the subsediment volume suggest that the collected 3-d data might not always support target detection, regardless of the algorithm employed. Factors such as sediment attenuation, interface scattering levels, the presence of gas bubbles in the sediment, and the relationship between sensor resolution and target dimensions mean that the upper limit on detection capability will likely be less than unity. To assess this possibility, the data from each target opportunity was first visually examined and rated in terms of anomaly size (large or small) and strength (strong or weak) in the imagery. Anomalies that were deemed both small and weak represent a “gray zone” in which detection may or may not actually be feasible.

With these human assessments as a backdrop, the performance of the proposed target detection algorithm at eight distinct data collections, delineated by location and time, are shown in Figure ES-2 for proud and buried targets. As can be seen from the figure, performance varies considerably across location (cf. collection letters), but also across time (cf. collection numbers), the latter variation suggesting strong environmental dependence (e.g., water temperature, microbial activity). However, in all cases, the automated detection performance comported with the expected range based on visual inspection of the imagery.
Figure ES-2. Performance of the Detection Algorithm for Each SVSS Data Collection for (a) Proud Man-made Targets and (b) Buried Man-made Targets, along with the Distribution of Visual Human Assessment Ratings.

Above each bar are the numbers of targets detected vice opportunities, and in brackets the range of targets deemed detectable based on visual human assessment.

Localized alarm data cubes (displayed as MIPs) extracted from Figure ES-1 of four targets are shown in Figure ES-3, along with corresponding object photographs taken during installation. (The 3-d alarm data cubes like in Figures ES-3(a)-(d) are what would be the inputs to the subsequent CNN classifier.)
Figure ES-3. (a)-(d) SVSS Alarm Cubes (Each Displayed As a Trio of MIPs) of Four Targets Extracted from Figure ES-1, and (e)-(h) Photographs of the Objects During Installation (pre-Burial).

The objects, along with the human-assessments of the sonar imagery in parentheses, are: (a) 4:1 solid aluminum cylinder proud (Large/Strong), (b) 2:1 solid aluminum cylinder buried 5 cm (Large/Weak), (c) 4:1 solid aluminum cylinder buried 3 cm (Large/Strong), (d) 10.2 cm diameter steel shot put buried 19 cm (Small/Strong); the cylinders have 15.2 cm diameters.

For the MuST system, the project team possessed neither ground-truth information nor sufficient amounts of data to make statistically significant detection-performance assessments. But applying the detection algorithm to the modest amounts of data the project team did have resulted in seemingly reasonable alarms. An example set of such alarms is shown in Figure ES-4.

Figure ES-4. Example Alarms Generated by the Detection Algorithm when Applied to a MuST Scene.

The combined detection performance from pooling the proud and buried targets of the various SVSS data collections is shown in Figure ES-5; a “score” based on the geometric mean of a size feature and an intensity feature is used to order the alarms. Performance is displayed in terms of
receiver operating characteristic-like (ROC-like) curves, where the probability of false alarm is replaced by false alarm rate (per unit volume). It should be noted that these results are for a considerable span of object sizes, some of which are even less than the sensor resolution.

Figure ES-5. Overall Performance of the Detection Algorithm for Each SVSS Data Collection.

4.3 CLASSIFICATION

The main interest in this section is to examine the ability of the CNN-based approach to successfully perform classification. Therefore, when presenting performance, the experiments assume that all targets were successfully flagged in the detection stage. Thus, the maximum possible area under the (ROC) curve (AUC) [10], a scalar summary measure of performance, is always unity. (A perfect classifier would have an AUC of unity.) To obtain the overall probability of both successful detection and correct classification, the vertical axis of the ROC-like curves would need to be scaled by the inverse of the total number of targets present in the scene imagery.

Performance is presented in the form of full ROC-like curves, with the abscissa corresponding to the more informative false alarm rate instead of the probability of false alarm. The probability of false alarm is the probability of incorrectly classifying clutter as a target; the false alarm rate is the number of such incorrect classifications per image area or volume. When considering only proud objects, the false alarm rate is given per image (seafloor) area; when considering only buried objects, the false alarm rate is given per image volume.

Performance of the eight trained CNNs, as well as the ensemble, is shown in terms of ROC-like curves in Figure ES-6. The AUC (for the corresponding ROC curve) is also provided in the legend. To provide a baseline measure of performance, the performance of the 3-d detection algorithm is also shown. While the full curves are informative, in practice, one must select a single operating point at which to make predictions. Therefore, on each CNN curve, the operating point corresponding to the natural decision threshold of \( \tau = 0.5 \) is also marked.
As can be observed from Figure ES-6, the 3-d CNNs greatly outperform the simple baseline detector, as would be expected. But more interestingly, the use of the ensemble of networks proves beneficial and removes the necessity of selecting a single best CNN architecture to employ. The complementary nature of the CNNs, and the unique clues that each uncovers and exploits to make predictions, leads to reduced false alarm rates. As a result, the ensemble approach can directly translate into cost savings during UXO remediation efforts.

![Figure ES-6](image)

**Figure ES-6. Classification Performance on the SVSS Test Data Set in Terms of ROC-like Curves for (a) Only Proud Objects and (b) Only Buried Objects.**

*The operating point for a $\tau = 0.5$ threshold is marked with a circle.*

It is also worth noting that these classification results are based on using only a single data representation, namely the 3-d imagery, as input to the CNNs. A worthwhile avenue to explore is the use of alternative data representations (e.g., acoustic color) in which complementary discriminative clues would be made more accessible to the CNN.

### 4.4 LIMITED-SCOPE CLASSIFICATION

Explainability of classifier predictions can be a useful tool to secure human trust of an algorithm’s decision-making process. With an eye toward that long-term goal, experiments were conducted to assess the feasibility of employing CNNs with *simulated* multi-representation acoustic-color sonar data for discriminating air-filled objects from water-filled objects.

CNNs were trained for seven different combinations of input data representations, in terms of frequency band and component (i.e., magnitude or phase). The performance of the CNNs for the different input data representations is presented in Figure ES-7. Additionally, ensembles that leverage different combinations of the CNNs (by averaging their individual predictions) are also considered. The AUC of each case is shown in the legend.

In Figure ES-7(a), it can be seen that a CNN trained on any of the representations was able to successfully discriminate the air-filled munitions from the water-filled munitions. However, in this case, the training data and test data – although disjoint – all corresponded to the same object, namely 155 mm munitions. As a result, the features that the CNN learned to rely on when discriminating
interior fill might be tied to this specific object. A stronger test of CNN generalization ability is shown in Figure ES-7(b), where the test objects are 105 mm munitions. In this figure, it can be seen that the CNN trained using broadband magnitude acoustic-color data (the red curve) was still able to reliably classify the test objects’ fill. This preliminary result suggests that this CNN indeed leverages attributes associated with the object’s interior fill, and more importantly, that these clues are ostensibly present in objects other than the specific type used for training.

Figure ES-7. For Different Input Data Representations, Classification Performance for Discriminating Air-filled and Water-filled Objects Using 155 mm Munitions as Training Data, and then (a) Testing on Other 155 mm Munitions or (b) Testing on 105 mm Munitions.

(Note the logarithmic horizontal-axis in (a).)
5.0 IMPLICATIONS FOR FUTURE RESEARCH AND BENEFITS

The work summarized in this report covers only one year of an envisioned four-year project that ended prematurely (due to an organization change of the principal investigator). Nevertheless, the progress made during this abbreviated period provides a solid foundation from which to further this line of research. Because the algorithms were purposely developed to be functional with measured data from existing systems, they should be readily deployable in a short time frame for use in actual remediation efforts. This result can be achieved by executing the remainder of the original project plan, which includes rigorous testing at new SERDP UXO test-bed sites.

6.0 REFERENCES


