Sonar Image Composition for Semantic Segmentation Using Machine Learning

William Ard
Louisiana State University
Baton Rouge, LA
ward2@lsu.edu

Corina Barbalata
Louisiana State University
Baton Rouge, LA
cbarbalata@lsu.edu

Abstract

This paper presents an approach for merging side scan sonar data and bathymetry information for the benefit of improved automatic shipwreck identification. The steps to combine a raw side-scan sonar image with a 2D relief map into a new composite RGB image are presented in detail, and a supervised image segmentation approach via the U-Net architecture is implemented to identify shipwrecks. To validate the effectiveness of the approach, two datasets were created from shipwreck surveys: one using side-scan only, and one using the new composite RGB images. The U-Net model was trained and tested on each dataset, and the results were compared. The test results show a mean accuracy which is 15% *higher for the case where the RGB composition is used when compared with the model trained and tested with the side-scan sonar only dataset. Furthermore, the mean intersection over union (IoU) shows an increase of* 9.5% *using the RGB composition model.*

1. Introduction

The Thunder Bay National Marine Sanctuary is a part of the United States' National Marine Sanctuaries. Located on Lake Huron's Thunder Bay, it protects a collection of more than 100 known shipwrecks. Wooden-wheelers, steel-hulled steamers, and more unique vessels can be found in this sanctuary, showcasing the importance of the Great Lakes' commercial shipping industry over the past two hundred years [2]. While many of the shipwrecks have been accurately located and identified, archaeologists suspect that another 100 shipwrecks are yet to be identified within the Thunder Bay National Marine Sanctuary [1].

The localization and study of shipwrecks offers a gateway towards a better understanding of history, cultural heritage, and maritime human activities. It also demonstrates how rough environments and human error can lead to loss of life in marine environments. Furthermore, accurate mapping of areas of interest can aid management efforts and reduce the risk of misappropriation [4]. The study of submerged artifacts and underwater sites is possible through the merging of expertise in geography, oceanography, engineering, and computer science. For obtaining a detailed understanding of archaeological sites, data collection is done using either divers, ships, or more recently marine vehicles such as Remotely Operated Vehicles (ROVs) or Autonomous Underwater Vehicles (AUVs).

Sonar devices represent some of the most common sensors used for seafloor surveys [19] as they can be either deployed on board marine vehicles or attached to ships covering large areas. One of the most common active sonar types used for detecting objects underwater is the side scan sonar. Such data sets can be characterized by shadows, low resolution, and lack of texture, limiting the capabilities of automatic target recognition (ATR) algorithms, especially when the targets are shipwrecks. Furthermore, side scan sonar devices cannot measure bathymetry, so to provide comprehensive maps of the floor, they can be used with other depth-measuring senors such as multibeam or single beam sonar. Nevertheless, the correlation of all these data sets for automated surveys can be challenging. With the advancement of technology that allows for the integration of swath bathymetry and dual-frequency side scan sonar systems, it is possible to have access to both sonar imagery and bathymetry data. This offers additional information that can be used to better understand the environment and identify areas of interest.

Once data is collected, marine archaeologists interested in shipwrecks will analyze it and look for areas of interest, but this type of manual annotation of large datasets can be tedious and time consuming [19]. With advancements in computer vision and machine learning approaches, semi-automated and automated labeling methods have been developed for the identification of archaeological sites [13], and more recently for the detection of underwater archaeological sites [4]. When sonar datasets are used in automated identification of shipwrecks, the focus has been on the utilization of sonar imagery. Nevertheless, [6] highlights the relevance of bathymetry information for shipwreck detection. Furthermore in [14] bathymetry informa-

tion and acoustic backscatter mosaics are used together for seafloor segmentation, looking at the identification of valleys and ridges through the use of classical computer vision techniques.

Driven by these findings that demonstrate the benefits of correlating sonar imagery with bathymetry data, this paper proposes an approach for consistently merging all available information from sonar systems, creating a new type of dataset. Furthermore, this paper will demonstrate the benefits of these datasets for machine learning segmentation, even in the case where limited training sets are available. The output of this research is an automated pipeline for shipwreck localization based on sidescan sonar and bathymetry information.

The rest of the paper is structured as follows: in Section 2 an overview of relevant work in the area of machine learning and automated segmentation for shipwreck identification is presented; in Section 3 the proposed methodology is presented; in Section 4 the performance of the proposed approach is evaluated, and in Section 5 the papers concludes with avenues for future work.

2. Background

The classical ATR process consists of pre-processing, region of interest (ROI) identification, region segmentation, feature extraction, and class identification (classification) [19]. Other approaches have focused directly on semantic image segmentation of sonar data sets [25, 17]. This has also been demonstrated in fields such as medical diagnostics [21], satellite imaging analysis [26], autonomous driving [10], and facial recognition [15]. Semantic image segmentation is the process of labeling regions of an image. Unlike a traditional image classification problem where the whole image has a single output label, semantic segmentation involves predicting the class of each pixel in an image. Research focused on the segmentation stage in sonar data sets have used methods such as active contours [11], clustering [3], Markov random field [16] or convolution neural network (CNN) [27]. In the past years machine learning approaches have been central in sonar segmentation.

In [23], the authors are aiming to overcome over-fitting problems and improve accuracy by merging deep neural networks with transfer learning. This approach also enables accurate segmentation when a limited amount of data is available for the training stage. This is done by using boat and non-ship images to generate artificial shipwreck sonar images. Another transfer learning approach that addresses the limited labeled shipwreck datasets is presented in [9]. This approach uses a pre-trained YOLO detector to learn pixel intensities and provides a solution for anomaly detection. YOLOv5 is used with a transformer module for real time object detection in [28]. The transformer module is added to ensure the target is the focus of the detection and ignore features in the proximity of the target. This enables an attention mechanism in the architecture, handling the limited texture and features in sonar images. A deep active instance segmentation method combining Region-Based Convolution Neural Network (R-CNN), active learning, and balanced sampling is presented in [22]. The approach has benefits when a limited amount of labeled samples are available, leading to better results for underwater shipwreck detection compared with a classical R-CNN.

In [4] the authors provide an investigation into the viability of swath bathymetric and topographic LiDAR data for ATR, and specifically shipwreck detection. The paper shows that shipwrecks can be detected in various levels of turbid waters using the YOLOv3 algorithm. Furthermore, [14] shows the benefits of leveraging both bathymetry data and acoustic backscatter to produce a segmentation map of the sea floor. As opposed to the previous approaches discussed in this background section, the algorithms used to produce the segmentation are texture classification, pattern recognition and morphological opening.

One of the recent trends in machine learning approaches for semantic segmentation is to use the U-Net architecture, that was introduced in 2015 [18]. It is a full CNN and is designed to work with fewer training images. It consists of symmetrical contracting and expanding paths. The contracting, or downsampling, path is much like a traditional CNN, and at each downsampling step the number of feature channels is doubled. During the expanding path, the feature map is upsampled and concatenated with the feature maps from each downsampling step via skip connections. The skip connections from earlier layers allow for the reconstruction of finer details around the segmentation boundaries, while the contracting layers give the context.

U-Net was originally used for biomedical image segmentation in cases where the number of available annotated samples can be limited. Data augmentation was achieved via elastic deformations of available training images. This is particularly useful in the case of biomedical imaging, as tissue deformations are realistically simulated. The benefits of U-Net for side-scan sonar data are presented in [24], where a 1D architecture is used for seabed detection from backscatter data.

This paper aims to leverage the U-Net architecture for shipwreck identification in a new type of dataset generated from bathymetry and side scan imagery. The details of the proposed approach are presented in the following section.

3. Methodology

This section discusses the sonar and bathymetry integration into a new imagery format, and the implementation of the U-Net algorithm for high accuracy shipwreck identification.

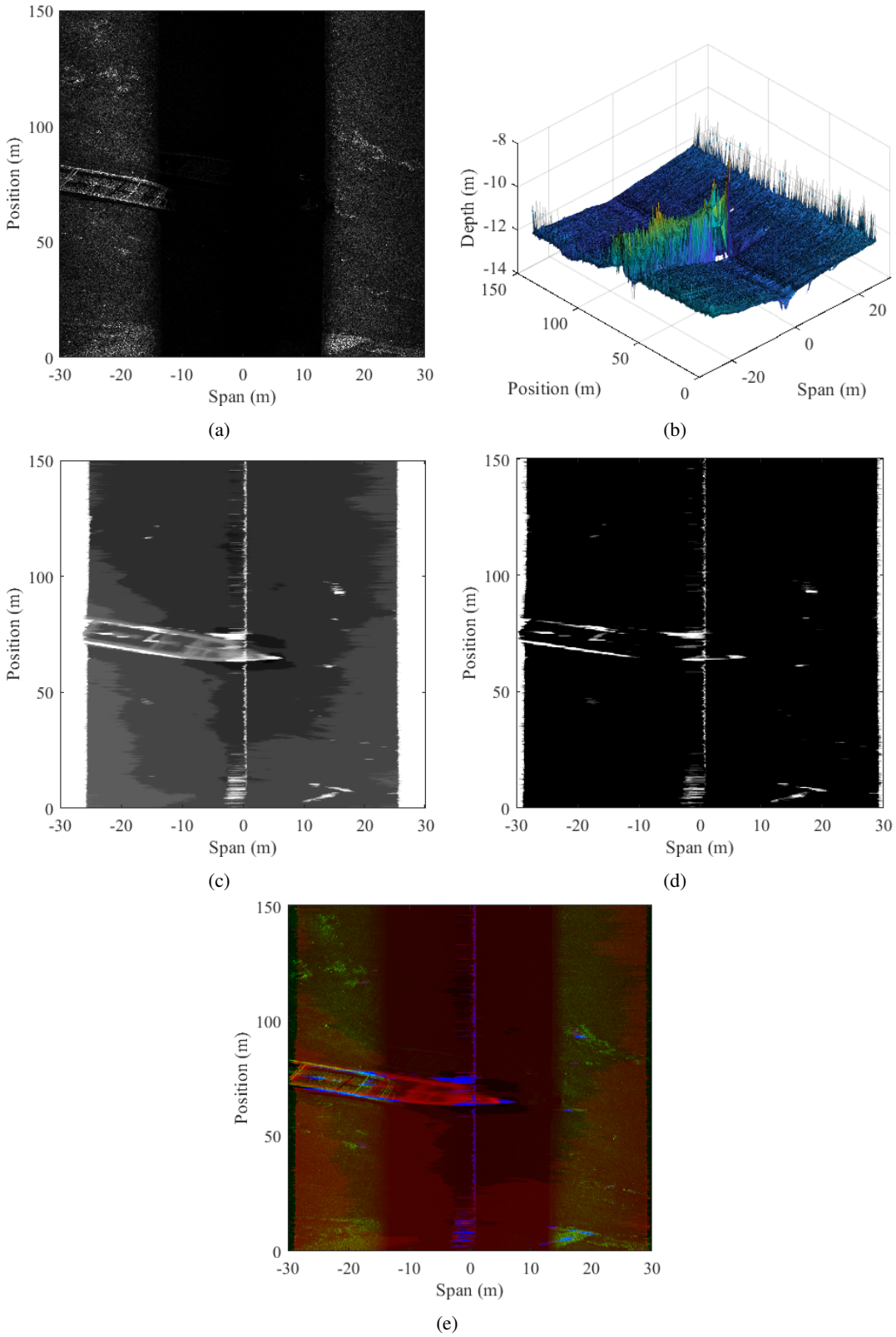


Figure 1: D.M. Wilson, wrecked 1894, Thunder Bay National Marine Sanctuary: (a) Side-scan sonar image, (b) 3D relief map, (c) 2D relief map, (d) bathymetry quality measurements, (e) image composition from (a), (c) and (d) images.

3.1. Data collection and processing

Side scan sonar and bathymetry data were collected using an EdgeTech 2205 sonar system [7] integrated with an L3Harris Iver3 AUV [12] at the Thunder Bay National Marine Sanctuary.

This system is able to collect sidescan sonar data, bathymetry information, pitch-roll data, GPS coordinates, altitude, pressure, and heading. Survey data is written in EdgeTech's native JSF binary file format [8]. A single sonar scan stores two messages per ping, one for port and one for starboard. By concatenating each sonar message recorded during a mission, a side-scan sonar image of the seafloor can be generated. A sample sonar image is shown in Figure 1a. Sonar data is prone to speckle noise that is a result of interference of the backscattered waves.

Bathymetry messages, like sonar messages, are separated for the port and starboard side. Each bathymetry message contains the time delay, angle, and amplitude of each assumed seafloor echo. There is also a quality factor, in which EdgeTech's processing algorithm uses a filter to indicate points whose phase differences are greater than a specified tolerance. The quality factor is an indication of how well interstave phase measurements agree. When measurements do not agree this is typically due to high noise or multipath interference. Multipath interference occurs when the sonar reflection returns via multiple routes. This is typically due to the scattering of the reflections by shipwreck features. The reflections that return are out of phase, and the signal is attenuated.

A relief map of the seafloor is constructed from bathymetry data. Using the time delay and the speed of sound in water reported by the sonar system, the nominal slant range, SR, is defined as:

$$SR = \frac{c}{2} \times t_e \tag{1}$$

where c is the sound velocity and t_e is the echo time. Using the angle from nadir, α and the slant range, the below sounder depth, z, and distance x along the span are calculated with Equation (2) and Equation (3), respectively.

$$z = SR \times \cos(\alpha) \tag{2}$$

$$x = SR \times \sin(\alpha) \tag{3}$$

By concatenating each bathymetry message, a relief map of the seafloor is created. A sample 3D map is shown in Figure 1b and a 2D map of the same bathymetry data is shown in Figure 1c.

Once the sonar and bathymetry information have been processed, the next step is to create a new multi-channel image, simulating an RGB configuration. The new image is created via concatenation along the third dimension. Sidescan sonar data is stored in the green channel of the image, and the red channel represents the 2D relief map. Since areas where bathymetry readings are poor can be indicative of shipwreck features, this information is preserved in the blue channel, as shown in Figure 1d. An example of the new generated multi-channel image is shown in Figure 1e. Using these steps a new set of images have been generated that align high quality bathymetry information with sonar imagery.

3.2. Image segmentation

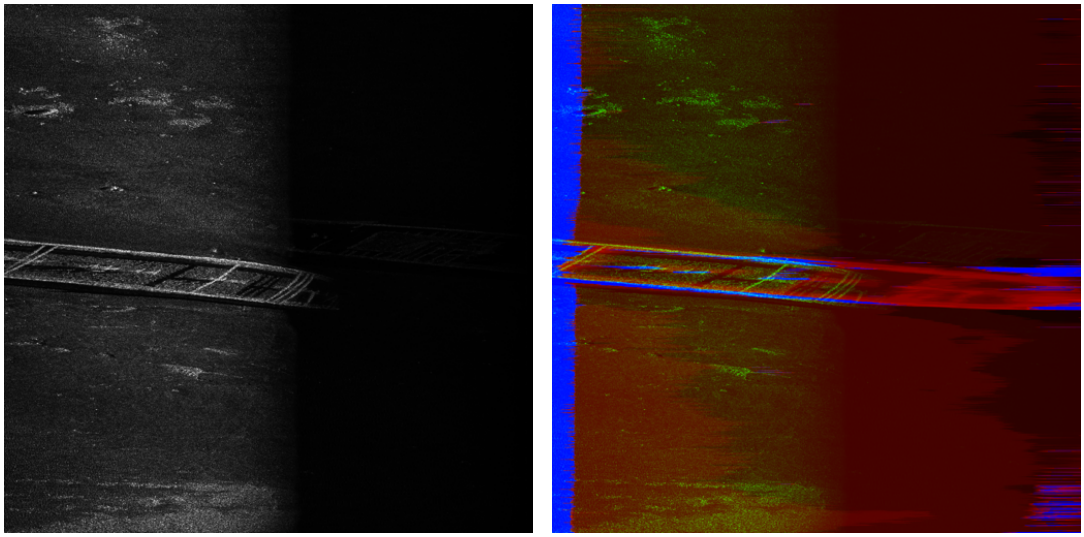
To validate the benefits of the new muli-channel dataset for automatic shipwreck identification, a machine learning approach has been implemented. A U-Net network with an encoder-decoder depth of 3 was used. Models were trained using the Adam (adaptive moment estimation) optimizer over 10 epochs with a learning rate of $1e-4$. A randomized $80-10-10$ training-validation-test split was used on the labeled data set consisting of 252 images, each with dimensions 256×256. Training images were taken from the separate port and starboard side readings of the sounder. Splitting the readings in this manner is beneficial, as it reduces the amount of memory required during training. The U-Net model was trained and evaluated for two cases. The first case used a dataset of sonar images, and the second case used a dataset composed of multi-channel images including sonar and bathymetry data. Figure 2a shows a sample training image used in the sonar only version of the model, and Figure 2b shows a sample training image using the multi-channel version.

Datasets were manually labeled using the LabelMe annotation tool [20]. Data manipulation and network implementation were done using Matlab.

4. Results

Using the randomized $80-10-10$ training-validation-test split, of the 252 images used in each dataset, 202 were used for training, 25 were used for validation during training, and 25 were used for testing. Ten epochs took approximately 45 minutes to train on a NVIDIA GeForce RTX 3080Ti GPU.

Qualitative evaluation of the proposed pipeline can be seen from Figure 3. In this example three different surveys are presented. The first column presents the manually labeled images, the second column presents the results of the U-Net segmentation when the new image composition is used for training, and the third column presents the results when only the side-scan sonar images were used. It can be seen that although using the side-scan sonar the contour area is correctly identified, the overall coverage is less compared to the case when the image composition is used.



(a) (b)

Figure 2: Sample training image of the D.M. Wilson, taken from the port side of the sounder: (a) side-scan sonar only image and (b) combined sonar and bathymetry image

The quantitative evaluation was done using all test images. Three metrics were used to evaluate the proposed architecture: mean accuracy, mean intersection over union (IoU), and mean boundary F1 (BF) score. Mean accuracy is the average pixel accuracy score of all classes in the image. Accuracy is defined as:

$$Accuracy = \frac{TP}{TP + FN} \tag{4}$$

where TP represents the number of true positives and FN represents the number of false negatives. Intersection over union is defined as the ratio of correctly classified pixels to the total number of ground truth predicted pixels:

$$IoU = \frac{TP}{TP + FP + FN} \tag{5}$$

where FP is the number of false positive pixels. BF score is the measure of how the predicted boundary matches the ground truth boundary, computed as presented in [5].

Table 1 gives a comparison of the performance of the U-Net architecture when sonar image composition dataset are used versus the side-scan sonar dataset only. The mean BF score validates the qualitative results, highlighting that the contours of the shipwreck are more clearly identified using only side-scan sonar images. Nevertheless, the overall area identified correctly as shipwrecks is significantly improved when the sonar image composition dataset is used. This can be seen through the mean accuracy and mean IoU metrics. The image composition model gives a mean accuracy which is 15% higher than that of the sonar only model, and a mean IoU that is 9.5% greater.

Model	Mean Accuracy	Mean IoU	Mean BF Score
RGB	0.7061	0.6254	0.5101
Sonar Only	0.5510	0.5320	0.5995

Table 1: Test results comparing U-Net predictions using image composition (RGB) model, versus U-Net with only side-scan sonar images.

5. Conclusion and Future Work

The paper introduced an approach to combine sidescan sonar data with bathymetry information into a single combined image. The benefits of this new dataset have been evaluated for automatic shipwreck identification using a supervised semantic segmentation approach, the U-Net architecture. Due to the feature enhancement capabilities of this method, the correctly identified areas have significantly increased compared to the case when only side-scan sonar images have been used. However, the precision of boundary identification of the shipwreck has decreased. Future work will focus on increasing the test set size, performing data augmentation, implementing additional semantic segmentation approaches, and implementing pre-processing strategies.

Acknowledgements

This project has been partially supported by NOAA award NA21OAR0110196 and NSF award #2122068. The authors would like to thank to Dr. Guy Meadows (Michigan

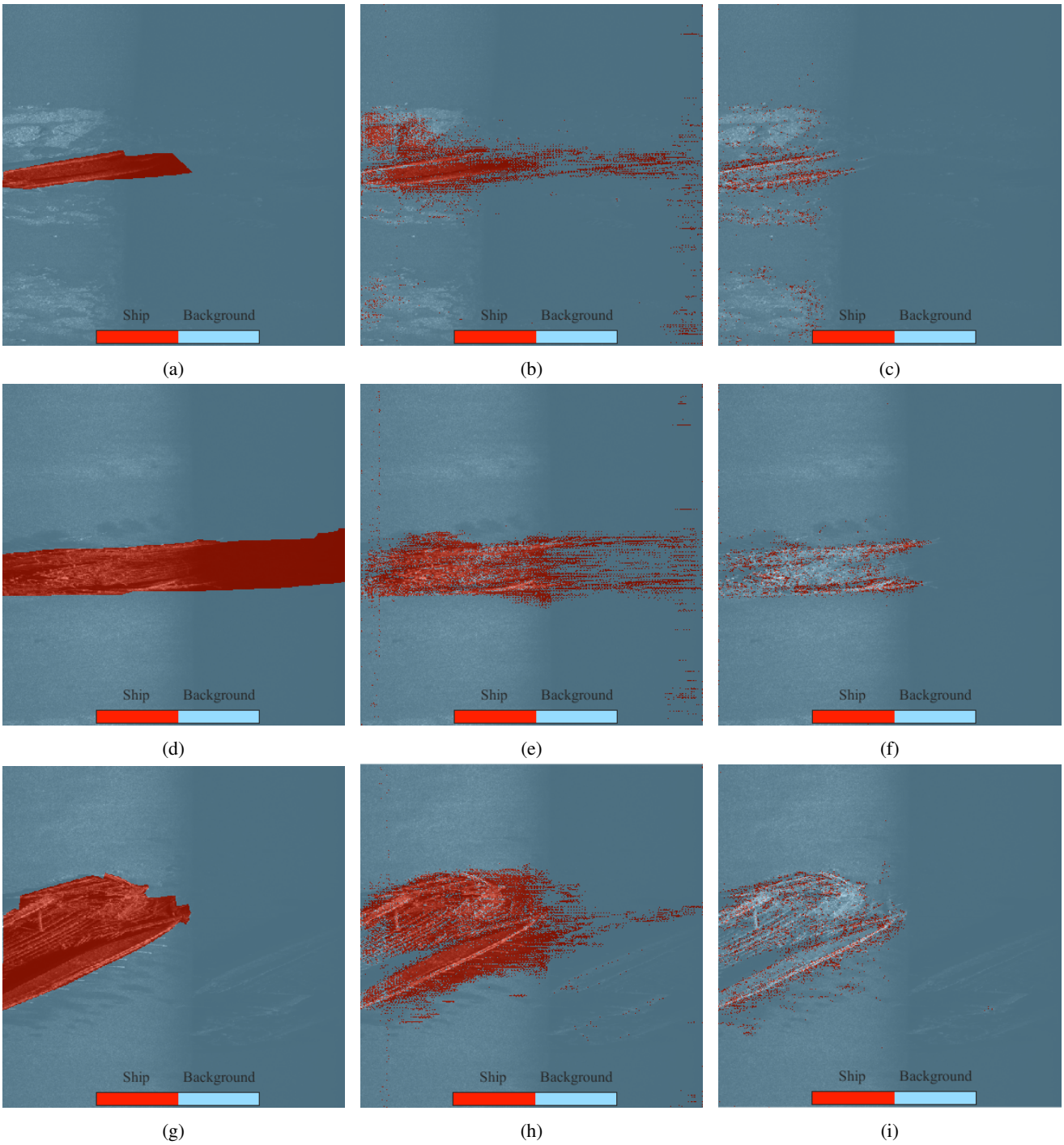


Figure 3: Semantic segmentation results: the first column (images a, d, g) represent the ground truth manually labelled, the second column (images b, e, h) present the segmentation results when the image composition was used, and the last column (images c, f, i) are the segmentation results when only side-scan sonar images were used.

Technological University) and Dr. Kathrine Skinner (University of Michigan) for the data collection and sharing.

References

[1] Holly A Bamford, Jefferson J Gray, Russ Green, Ellen Brody, John Wagner, Helene Scalliet, and Vicki Wedell.

Thunder bay national marine sanctuary boundary expansion: final environmental impact statement. 2014.

[2] Leah R Burgin. Managing michigan's underwater heritage: The past, present, and future of thunder bay national marine sanctuary'. *Ann Arbor*, 1001:48109–1354, 2015.

[3] Turgay Celik and Tardi Tjahjadi. A novel method for sidescan sonar image segmentation. *IEEE Journal of Oceanic Engineering*, 36(2):186–194, 2011.

[4] Leila Character, Agustin Ortiz Jr, Tim Beach, and Sheryl Luzzadder-Beach. Archaeologic machine learning for shipwreck detection using lidar and sonar. *Remote Sensing*, 13(9):1759, 2021.

[5] Gabriela Csurka, Diane Larlus, and Florent Perronnin. What is a good evaluation measure for semantic segmentation? In *Procedings of the British Machine Vision Conference 2013*, pages 32.1–32.11, Bristol, 2013. British Machine Vision Association.

[6] Dylan S Davis, Danielle C Buffa, and Amy C Wrobleski. Assessing the utility of open-access bathymetric data for shipwreck detection in the united states. *Heritage*, 3(2):364–383, 2020.

[7] EdgeTech. Sonar 2205. `https://www.edgetech.com/product/2200-and-2205-auv-uuv-rov-asv-usv-sonars/`.

[8] EdgeTech. Jsf file and message descriptions. `https://www.edgetech.com/wp-content/uploads/2019/07/0023492_Rev_G.pdf`, 2022.

[9] Dylan Einsidler, Manhar Dhanak, and Pierre-Philippe Beaujean. A deep learning approach to target recognition in side-scan sonar imagery. In *OCEANS 2018 MTS/IEEE Charleston*, pages 1–4. IEEE, 2018.

[10] Di Feng, Christian Haase-Schütz, Lars Rosenbaum, Heinz Hertlein, Claudius Gläser, Fabian Timm, Werner Wiesbeck, and Klaus Dietmayer. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. *IEEE Transactions on Intelligent Transportation Systems*, 22(3):1341–1360, 2021.

[11] Daniel Köhntopp, Benjamin Lehmann, Dieter Kraus, and Andreas Birk. Segmentation and classification using active contours based superellipse fitting on side scan sonar images for marine demining. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3380–3387. IEEE, 2015.

[12] L3Harris. Iver3. `https://www.l3harris.com/all-capabilities/iver3-ep-open-system-auv`.

[13] Lei Luo, Xinyuan Wang, Huadong Guo, Rosa Lasaponara, Xin Zong, Nicola Masini, Guizhou Wang, Pilong Shi, Houcine Khatteli, Fulong Chen, et al. Airborne and spaceborne remote sensing for archaeological and cultural heritage applications: A review of the century (1907–2017). *Remote Sensing of Environment*, 232:111280, 2019.

[14] Giuseppe Masetti, Larry Alan Mayer, and Larry Guy Ward. A bathymetry-and reflectivity-based approach for seafloor segmentation. *Geosciences*, 8(1):14, 2018.

[15] Toshanlal Meenpal, Ashutosh Balakrishnan, and Amit Verma. Facial mask detection using semantic segmentation. In *2019 4th International Conference on Computing, Communications and Security (ICCCS)*, pages 1–5, 2019.

[16] Max Mignotte, Christophe Collet, Patrick Pérez, and Patrick Bouthemy. Three-class markovian segmentation of high-resolution sonar images. *Computer Vision and Image Understanding*, 76(3):191–204, 1999.

[17] Maryam Rahnemoonfar and Dugan Dobbs. Semantic segmentation of underwater sonar imagery with deep learning. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 9455–9458. IEEE, 2019.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. Technical Report arXiv:1505.04597, arXiv, May 2015. arXiv:1505.04597 [cs].

[19] Yannik Steiniger, Dieter Kraus, and Tobias Meisen. Survey on deep learning based computer vision for sonar imagery. *Engineering Applications of Artificial Intelligence*, 114:105157, 2022.

[20] Kentaro Wada. labelme: Image polygonal annotation with python. `https://github.com/wkentaro/labelme`, 2018.

[21] Yuchao Wu, Lan Lin, Jingxuan Wang, and Shuicai Wu. [application of semantic segmentation based on convolutional neural network in medical images]. *Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi*, 37(3):533—540, June 2020.

[22] Fangjin Xu, Jianxing Huang, Jie Wu, and Longyu Jiang. Active mask-box scoring r-cnn for sonar image instance segmentation. *Electronics*, 11(13):2048, 2022.

[23] Lixue Xu, Xiubo Wang, and Xudong Wang. Shipwrecks detection based on deep generation network and transfer learning with small amount of sonar images. In *2019 IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS)*, pages 638–643. IEEE, 2019.

[24] Jun Yan, Junxia Meng, and Jianhu Zhao. Bottom detection from backscatter data of conventional side scan sonars through 1d-unet. *Remote Sensing*, 13(5):1024, 2021.

[25] Dianyu Yang, Can Wang, Chensheng Cheng, Guang Pan, and Feihu Zhang. Semantic segmentation of side-scan sonar images with few samples. *Electronics*, 11(19):3002, 2022.

[26] Naisen Yang and Hong Tang. Semantic segmentation of satellite images: A deep learning approach integrated with geospatial hash codes. *Remote Sensing*, 13(14), 2021.

[27] Fei Yu, Bo He, Kaige Li, Tianhong Yan, Yue Shen, Qi Wang, and Meihan Wu. Side-scan sonar images segmentation for auv with recurrent residual convolutional neural network module and self-guidance module. *Applied Ocean Research*, 113:102608, 2021.

[28] Yongcan Yu, Jianhu Zhao, Quanhua Gong, Chao Huang, Gen Zheng, and Jinye Ma. Real-time underwater maritime object detection in side-scan sonar images based on transformer-yolov5. *Remote Sensing*, 13(18):3555, 2021.