BLENDgänger: Generating Synthetic MBES Data for Underwater UXO Perception Tasks

Amos Smith[*,§], Nael Jaber[†,§], Leif Christensen[†], Martin Atzmueller[*,‡]

[*]Semantic Information Systems Group (SIS), Osnabrück University, Osnabrück, Germany
[†]Robotics Innovation Center, German Research Center for AI (DFKI), Bremen, Germany
[‡]Cooperative and Autonomous Systems, German Research Center for AI (DFKI), Osnabrück, Germany

***Abstract*—Unexploded ordnance (UXO) and discarded munitions in coastal waters pose serious environmental and safety risks. Effective explosive ordnance disposal (EOD) relies on accurate detection and characterization of UXO, often performed with multibeam echosounder (MBES) surveys. In practice, however, MBES data are corrupted by outliers stemming from sensor errors, environmental conditions, and acoustic interference. Machine-learning solutions demand large volumes of precisely labeled training data, but such labels are costly and time-consuming to obtain. To overcome these challenges, we present BLENDgänger, a procedural data-generation framework built on the Blender platform. BLENDgänger synthesizes realistic MBES bathymetric point clouds with configurable noise profiles and ground-truth annotations, enabling the rapid assembly of large-scale datasets for both semantic segmentation and 3D object detection of underwater ordnance. We demonstrate that models trained exclusively on BLENDgänger data achieve strong performance when evaluated on independent ex-situ MBES measurements. These results show that synthetic datasets can effectively bootstrap machine-learning workflows for UXO perception and inspection, reducing reliance on laborious manual annotation.**

***Index Terms*—MBES sonar simulation, UXO perception, point cloud segmentation, 3D object detection**

I. Introduction

Coastal marine waters worldwide are burdened by contamination from unexploded ordnance (UXO) and deliberately discarded munitions [1], which may destabilize over time and release toxic substances into the environment [2]. The removal of these munitions, named explosive ordnance disposal (EOD), is carried out by experts who first locate the UXO and determine its condition. Artificial intelligence has recently been applied to underwater sonar data, such as from multibeam echosounders (MBES), to assist in the EOD process and to reduce potential risks to EOD personnel [3]. MBES profiler sensors represent an efficient and accurate technology for acquiring bathymetry and backscatter measurements, but real-world surveys often result in up to 25% percent outliers [4]. These outliers, caused by operator error, sensor malfunctions, algorithmic limitations, harsh environmental conditions (e.g., rough weather, air bubbles), and acoustic interference from marine life or hydrothermal vents [5], degrade data quality and must be rigorously filtered during post-processing. Moreover, machine-learning approaches typically require extensive labeled datasets, the preparation of which involves labor-intensive manual annotation.

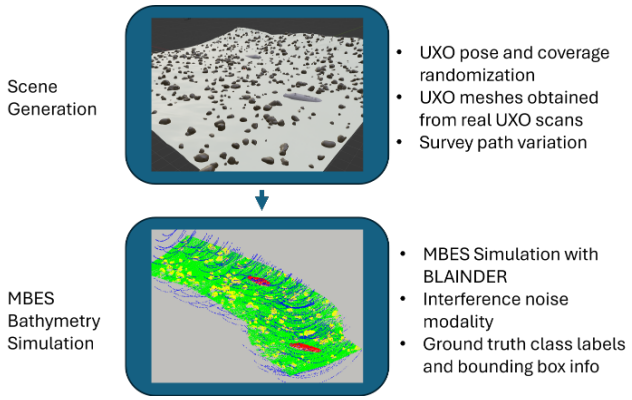

Fig. 1. Overview of the proposed BLENDgänger framework, which integrates scene generation and MBES bathymetry simulation to produce UXO-focused training datasets for perception models.

We introduce BLENDgänger, a Blender-based framework named after the German term *Blindgänger* (UXO), for the procedural generation of synthetic MBES datasets to support UXO perception. This framework, as shown in Fig. 1, is composed of scene generation and MBES bathymetry simulation and allows for the generation of labeled synthetic MBES datasets for the UXO use case. Leveraging BLENDgänger, we assemble a large-scale dataset to train machine learning models for both UXO segmentation and 3D detection. We then validate these models on an ex-situ dataset obtained from two distinct Autonomous Underwater Vehicle (AUV) platforms, with each AUV employing a different MBES sensor setup. Together, these experiments confirm that BLENDgänger's synthetic datasets can robustly support machine-learning workflows for underwater UXO perception, and could be extended in the future for other marine domain applications.

The main contributions of this work include:

- Generation of underwater scenes, including terrain, sensor trajectory, debris, and UXO object placement and

This work is jointly supported by the "HybrInt - Hybrid Intelligence through Interpretable AI in Machine Perception and Interaction" project (Zukunft Nds, Niedersächsisches Ministerium für Wissenschaft, Grant ID: ZN4219), the "CleanSeas" project (funded by BMFTR, Grant ID: 01IW22003), and the "CIAM" project (funded by BMWK, Grant ID: 03SX540D).
[§]These authors contributed equally to this work.

randomization.
- A corpus of UXO mesh objects to be used in simulation, created from 3D scans of real UXO objects.
- An interference noise modality model to mimic interference artifacts in MBES bathymetry data.
- Evaluation of UXO segmentation and 3D detection on real datasets, with the models being trained purely on synthetic data.
- Upon publication, the source code for BLENDgänger, including real dataset samples and simulation configurations used in this work, will be made available online [6].

The remainder of this paper is organized as follows. Section II reviews related work on UXO perception and MBES simulation. Section III details the methodology underlying the BLENDgänger framework, including scene generation and MBES simulation. Section IV describes the datasets used in this study, encompassing both real and simulated data. Section V presents the experimental setup and results of the evaluated perception models. Finally, Section VI discusses the experimental results and potential future research directions.

II. Related Works

A. UXO Perception

While sonar imagery (e.g., side-scan or forward-looking sonar) has previously been used in UXO detection, a growing body of research explores using MBES point clouds and bathymetric grids for object perception. Unlike intensity-based sonar imagery, MBES profilers provide 3D depth measurements that enable shape- and topology-based detection. In [7], a supervised machine learning model was trained using terrain derivatives such as slope and roughness obtained from MBES data. This model was used to distinguish potential UXO candidates from natural clutter in gridded depth data. Similarly, [3] used a UNet-style CNN to segment MBES bathymetric grids into UXO vs. background, showing that elevation-only input can be sufficient for initial UXO detection in seabeds. These geometry-based methods are more robust to reflectivity variations and offer geospatial accuracy, making them particularly useful for guiding autonomous inspection tasks.

Although still emerging, there is a growing interest in directly applying 3D point-based neural networks [8]–[12] to MBES bathymetric point clouds for underwater object detection. However, research in this area remains relatively limited. Himri et al. [13] conducted one of the early studies on underwater point cloud recognition by evaluating seven global descriptors from the Point Cloud Library (PCL) [14], examining their performance under different conditions including partial and global views, resolution levels, and added noise. Their findings showed that descriptor performance improves with global views and high-resolution data, but is significantly impacted by noise. In follow-up work, the same group used a local descriptor in combination with a Bayesian estimation model to segment underwater pipe components, such as connectors, valves, elbows, and R-Tee joints, from uncolored point clouds acquired by a laser scanner mounted on an AUV. Their best performance was achieved using the Clustered Viewpoint Feature Histogram (CVFH) descriptor [15].

Other studies have explored learning-based approaches. For instance, Martin et al. [16] collected RGB-colored point clouds using a binocular camera and trained the PointNet network [12] to segment pipes and valves. Similarly, Wang et al. [17] generated point clouds from stereo images and applied the Yolo V3 model [18] for object detection in optical images, projecting detections onto the point clouds to assist segmentation. Hu et al. [19] proposed a system combining laser scanning and binocular imaging to detect underwater pipeline systems. Their approach demonstrated high accuracy in extracting pipeline points and keypoint estimation using high-density laser-acquired point clouds, focusing specifically on pipeline identification.

Another study using the BV5000 3D acoustic sensor explored underwater detection of abandoned tires through 2D and 3D deep learning approaches. Pre-processed point clouds were transformed into bird's eye view images for 2D detection using Faster R-CNN and YOLOv3, and segmented for 3D classification with PointNet and PointConv. While both methods achieved high accuracy in detecting and classifying tires, the 3D point clouds used were extremely dense; similar in quality to LiDAR data; which may not reflect the lower resolution and sparser nature of typical multibeam sonar point clouds encountered in broader underwater applications.

All the aforementioned works rely on high-resolution sensing technologies such as RGB-colored point clouds, laser scanners, or 3D mechanical scanning sonars (MSS). While effective in clear environments, these sensors are limited by light-dependent operation, slow scanning rates, high costs, or limited coverage, making them less practical for large-area mapping. In contrast, MBES profilers offer wide coverage and real-time acquisition but produces sparser and noisier data, with greater challenges for object detection and classification.

Limited work addressing object recognition in sparse MBES maps is represented by [20], where a SECOND [10] 3D object detector was trained on MBES point clouds to automatically detect boulders, illustrating how similar techniques could be adapted for UXO detection. Another recent work [21] utilized a transformation-equivariant model for detecting 14 different underwater objects (including UXOs) in MBES data and demonstrated successful performance on real-world MBES scans. This model is able to localize objects regardless of their orientation or position, a critical requirement in underwater environments where object poses can vary unpredictably. In [22], segmentation of underwater pipeline MBES data is performed by converting MBES point clouds into "waterfall" images and 2D projections, and applying a UNet-based 2D image segmentation approach. The performance of these 2D projection-based methods was compared with that of a 3D point-based segmentation approach, with results showing that both approaches achieved comparable segmentation performance.

In the absence of large annotated field datasets, synthetic

MBES point cloud simulations generated using physics-based acoustic models can be used to train or pre-train point-based networks, helping overcome data scarcity while allowing for controlled variation in UXO type, orientation, and seabed condition.

B. MBES Point Cloud Simulation

A persistent challenge in this domain is the limited availability of large, annotated MBES point cloud datasets containing real UXO. To address this, researchers are increasingly turning to synthetic MBES simulation pipelines, where physics-based acoustic models generate realistic sonar returns from virtual seafloor scenes populated with UXO-like targets. These synthetic datasets can be used to pre-train point-based models and support controlled studies of object size, orientation, burial state, and environmental clutter. This approach offers a scalable path forward for training and benchmarking detection models in the absence of real-world field data.

Synthetic point cloud generation is commonly supported by simulation frameworks such as Gazebo [23], Helios++ [24], and Isaac Sim [25], which use raycasting to sample points on object surfaces within virtual 3D environments. These tools are widely used in robotics and perception research, as they can generate dense geometric point clouds along with semantic labels (e.g., object classes or instance IDs) for supervised learning tasks. However, while these point clouds are useful for general vision or LiDAR-based applications, they lack acoustic realism. Specifically, they do not simulate the physical behavior of sonar, such as beam directivity, backscatter intensity, reverberation, or multipath effects. As a result, the generated data does not reflect the distortions, intensity variations, or occlusion patterns inherent in real sonar returns, limiting their suitability for training or evaluating models in underwater acoustic sensing contexts.

Several open-source simulators now incorporate high-fidelity acoustic modeling to better replicate the physics of underwater sonar sensing. HoloOcean, developed on Unreal Engine, models profiling and imaging sonar using GPU-accelerated ray tracing and sonar equations that account for beam directivity, speckle noise, and multipath effects. It supports Python scripting, ROS integration, and custom 3D environments, making it one of the most versatile and physically grounded sonar simulators for underwater robotics [26].

OceanSim, a recent open-source framework based on NVIDIA's Omniverse and Isaac Sim, provides multi-sensor simulation, including visual and acoustic sensors, through physically-based rendering and sonar-inspired ray tracing. It outputs both synthetic sonar imagery and point clouds, with a focus on realism and compatibility with modern ML pipelines [27]. For ROS/Gazebo-based workflows, the DAVE multibeam plugin offers a CUDA-accelerated acoustic model, simulating echo intensities using point-scatter physics and beam noise. This plugin produces raw A-scan data that can be post-processed into sonar images or structured point clouds [28].

In contrast, tools like Stonefish provide more lightweight alternatives. While these simulators offer limited physical modeling, such as single-ray sonar beams or analytic echo generation, they are well-suited for rapid dataset generation or signal processing experiments [29], [30]. A particularly flexible tool for synthetic dataset creation is BLAINDER, a Blender-based framework designed to render labeled sonar and vision data in underwater scenes [31]. While BLAINDER lacks advanced acoustic physics features like beam pattern or multipath simulation, its integration with Blender enables users to fully generate, customize, and simulate underwater scenes and MBES point clouds entirely within a single tool. This self-contained approach streamlines the creation of annotated datasets for perception and AI tasks, including object detection and segmentation.

While these existing acoustic simulation tools have advanced the state of synthetic dataset creation for underwater perception tasks, most either lack the physical fidelity needed for MBES bathymetry data or do not provide seamless scene generation and annotation within a unified pipeline. BLENDgänger bridges this gap by combining a Blender-based procedural modeling workflow with a customizable BLAINDER-based MBES simulation, enabling the rapid creation of large, annotated datasets that realistically capture UXO, debris, and terrain variability as well as interference artifacts. This integrated approach not only supports machine learning research in UXO perception, but also provides an extensible platform for simulating a wide range of underwater sensing scenarios.

III. Methods

A. Scene Generation

Underwater environments are procedurally generated in Blender, an open-source 3D graphics suite. The generation

Fig. 2. Top-down view of UXO object meshes and corresponding real-life imagery. Left: meshes used in this work, from top to bottom: mortar shell, mine, artillery shell, deformed artillery shell, and 500 lb aircraft bomb; right: real-life images showing the rusted state of the corresponding UXO objects.

of simulated underwater scenes begins with the creation of a high-resolution digital terrain model using the A.N.T. Landscape tool. This tool provides an array of adjustable parameters to control terrain noise and depth profiles, enabling the synthesis of realistic seafloor topography. Once the terrain is generated, a sensor trajectory is defined at a user-specified height above the seabed, with the option to follow either a straight path or a path containing randomized heading deviations to mimic realistic survey motions. To further enhance scene realism, meshes representing small boulders are stochastically distributed over the terrain, with user-controlled parameters for boulder density and size. At this stage, user-created meshes representing objects of interest can also be inserted.

For this work, we include meshes derived from five real UXO objects. The meshes, shown in Fig. 2, were created from 3D scans performed with an Einscan H2 handheld scanner. The UXOs span a range of types, including a small 8 cm-diameter mortar shell as well as a large 500 lb general-purpose aircraft bomb (GP 500 LB MK3). Placement of UXOs in the scene is accomplished by sampling random terrain locations near the sensor trajectory to ensure adequate visibility, followed by the application of randomized orientations and vertical offsets. This approach produces scenarios featuring UXOs with varying degrees of coverage and exposure above the seafloor.

B. MBES Point Cloud Simulation

Once the environment is established, synthetic MBES bathymetry data is generated in the form of 3D point clouds with per-point semantic labels using the BLAINDER add-on [31]. BLAINDER utilizes the previously defined sensor trajectory to iterate through discrete positions, casting rays according to user-specified sensor parameters such as field of view and the number of beams per swath. Each valid ray–object intersection is recorded as a point in the simulated point cloud. A semantic class label is assigned to the point based on the corresponding object's class as defined during the scene generation.

The simulation provides precise control over additional parameters, including water-column acoustic velocity profiles and sensor noise characteristics. During raycasting, random noise is introduced to each beam, modeled as a combination of an offset and Gaussian noise along the direction of the ray. While this approach sufficiently emulates ideal MBES profiler operation, real-world MBES bathymetry datasets are often contaminated by structured noise due to sources such as operating machinery or side-lobe interference. Such artifacts often exhibit distinctive patterns, for example arc-shaped distortions that adversely affect range measurements and degrade the resulting point clouds.

To address this discrepancy, we extend BLAINDER's noise modeling capabilities with an interference noise modality. This mechanism is controlled by two user-defined parameters: `interference_chance_per_swath` and `interference_chance_per_beam`. The parameter `interference_chance_per_swath` specifies the probability that a given MBES swath will exhibit interference, while `interference_chance_per_beam` defines the probability that an individual beam within an affected swath is contaminated. If a beam is determined to be affected by interference, its range measurement is perturbed by sampling from a designated per-swath noise distribution; additionally, the semantic class label for that point is set to "noise". Fig. 3 illustrates simulated MBES point cloud data generated using

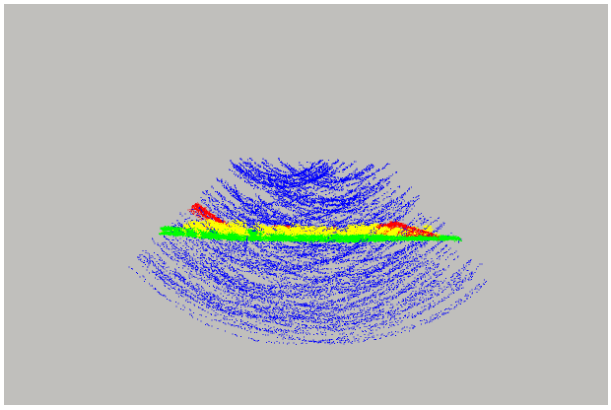

Fig. 3. Along-track view of a simulated MBES point cloud. Points affected by the proposed interference noise modality are labeled and displayed in blue, enabling development of models that learn to automatically reject such artifacts. Displayed are additional labels for the classes of seabed (green), boulders (yellow), and UXO (red).

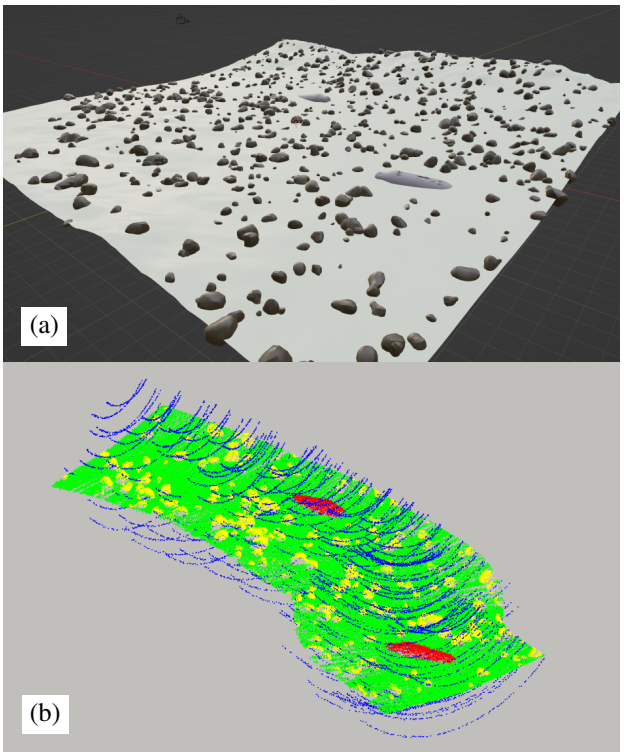

Fig. 4. Example result of underwater scene generation (a) and associated synthetic MBES bathymetry data (b) with labels corresponding to noise (blue), seabed (green), boulders/debris (yellow), and UXO (red).

this interference noise modality.

This interference noise model is intentionally lightweight and built on simplified assumptions, providing a practical means to emulate noise artifacts observed in real-world data. The parameters are configurable, enabling alignment with empirically observed noise patterns. By explicitly labeling interference artifacts, the model facilitates the development and training of perception algorithms that can learn to automatically detect and reject such noise during subsequent analysis.

Owing to the extensive parameterization of both environment creation and data-generation workflows, these processes can be executed iteratively to yield large volumes of diverse point clouds accompanied by corresponding ground-truth labels. Fig. 4 illustrates a representative output of the procedural environment generation and resulting labeled point cloud.

IV. Data

A. Real Data

To validate perception models trained on synthetic data, we collected real-world datasets in a controlled environment designed to preserve essential environmental variables. Data acquisition was performed using two distinct AUVs to assess the generalizability of synthetic data produced by BLENDgänger across platforms with differing characteristics. Specifically, we deployed the DeepLeng AUV [32], equipped with a R2Sonic 2020 MBES, and the Cuttlefish AUV [33], equipped with a Teledyne BlueView M1350 MBES. An overview of the MBES configurations for each vehicle is provided in Table I. The AUV's onboard pose estimator was used for accurate data georeferencing like any open-world mission.

To accurately replicate realistic terrain noise, the test site featured a pebble-covered seabed, and both genuine UXO objects and incidental debris were placed within the environment. Fig. 5 provides camera images of the setup and AUVs during a data acquisition run. The resulting MBES scans, an example of which is illustrated in Fig. 6, exhibit real-world noise and target characteristics while maintaining the advantages of a controlled setup with known ground truth.

From the collected data, six subsets were selected for evaluation. Two datasets from the Cuttlefish platform capture orbiting maneuvers around the test bed, while a third dataset was acquired during a straight-line survey over the test area. Three other datasets were obtained from DeepLeng performing linear passes at varying speeds. For each dataset, each MBES swath is transformed into a global frame by using inertial navigation system (INS) positioning information, similar to a real survey process. All datasets were manually annotated with segmentation labels and bounding boxes for use in evaluation of perception models trained on synthetic data.

TABLE I
MBES Sensor Comparison for DeepLeng and Cuttlefish AUVs

	DeepLeng	Cuttlefish
MBES Sensor	R2Sonic 2020	Teledyne BlueView M1350
Field of View	60°x1°	45°x1°
Beam Width	1°x1°	1°x1°
Number of Beams	256	256
Frequency	400 kHz	1.35 MHz
Max Range	200 m	30 m

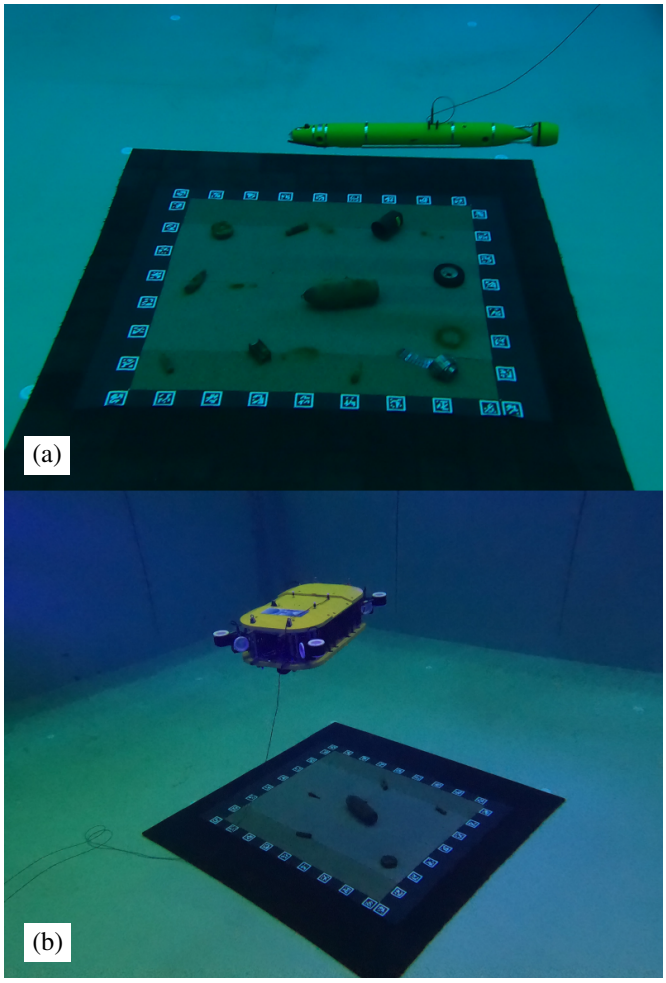

Fig. 5. Camera images of test bed setup and DeepLeng AUV (a) and Cuttlefish AUV (b) during data collection.

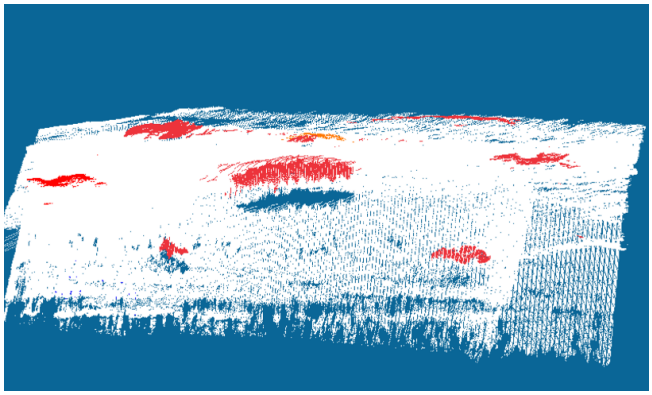

Fig. 6. Example MBES point cloud obtained from the DeepLeng AUV, with UXO and debris objects manually colored in red for visualization purposes.

B. Synthetic Data

A synthetic training dataset was generated using the proposed BLENDgänger framework. The dataset comprises 1000 scenes, each covering an area of 20 m $\times$ 20 m, with multiple instances of the 500 lb target UXO present in each scene. Diverse environments were achieved by randomly sampling sensor trajectories, terrain morphologies, and boulder distributions as described in Section III-A. The interference noise parameters, `interference_chance_per_swath` and `interference_chance_per_beam`, were set to 0.3 and 0.6, respectively, with the interference noise offsets uniformly sampled in the range of 2.5 m to 10 m. Sensor trajectory altitudes were sampled uniformly between 4 m and 7 m above the seafloor. A configuration file containing all parameters used for this dataset generation is available in the accompanying code repository.

V. Experiments and Results

The synthetic dataset was used to train models for two perception tasks: 3D semantic segmentation and 3D object detection, both targeting the identification and localization of the 500 lb UXO. This particular UXO was selected as the experimental focus due to its consistent and prominent visibility across all real datasets, enabling reliable evaluation and comparison of model performance.

A. 3D Semantic Segmentation

For 3D semantic segmentation, we implemented a sparse 3D U-Net using the Minkowski Engine [8], tackling point-wise classification into four categories: noise, seabed, debris, and UXO. We adopted the Res16UNet18 architecture, in which sparse 3D convolutions operate on voxelized point clouds with a resolution of 0.05 m. In this architecture, "16" denotes the base number of feature channels in the initial convolutional layer, and "18" indicates a ResNet-18–like depth, incorporating residual blocks within the U-Net framework for effective multi-scale feature learning.

Our synthetic dataset comprised 1,000 labeled point clouds, randomly split into 950 for training and 50 for validation. The network was trained from scratch using stochastic gradient descent and categorical cross-entropy loss, with a batch size of four point clouds. Data augmentation, such as center cropping, random rotations, and translations, was applied to improve generalization. Training proceeded for 25 epochs on an NVIDIA GeForce RTX 3080 Ti GPU. Model performance was evaluated using the Intersection over Union (IoU) metric, achieving a mean IoU of 85.5 on the validation set. The per-class IoU scores were 95.9 (noise), 98.4 (seabed), 87.9 (debris), and 60.0 (UXO).

The trained model was subsequently evaluated on the real datasets as detailed in Section IV-A, with IoU metrics computed separately for each platform to assess generalization across varying systems and MBES sensors. The DeepLeng dataset achieved a higher mean IoU of 0.475, compared to 0.336 for the Cuttlefish dataset. For the UXO class specifically, the IoU was 0.763 on DeepLeng data and 0.150 on Cuttlefish data. Complete per-class IoU scores for each dataset are provided in Table II. Due to a high concentration of noise in the Cuttlefish dataset, the "debris" class was not manually labeled and is therefore excluded from the Cuttlefish performance metrics. Example segmentation outputs for both the DeepLeng and Cuttlefish datasets, together with their ground truth counterparts, are displayed in Fig. 7 and Fig. 8, respectively. By treating noise as an explicit class, the network identifies spurious returns directly; these points can then be discarded, streamlining downstream EOD workflows and improving the fidelity of subsequent detection steps.

B. 3D Object Detection

For 3D detection, we adopted prior work of [21], which introduced a transformation-equivariant model for detecting various underwater objects in MBES data and demonstrated successful performance on real-world MBES scans. This model is able to localize objects regardless of their orientation or position, a critical requirement in underwater environments where object poses can vary unpredictably.

For the detection task, a subset of the dataset was used to train the aforementioned model. Specifically, 100 point clouds containing the 500 lbs UXO object were selected and split into 80% for training and 20% for validation. Since the transformation model is designed to learn generalizable object features that remain consistent despite variations in viewpoint or orientation, data augmentation was deemed unnecessary.

TABLE II
3D Segmentation IoU Metrics per AUV

IoU	noise	seabed	debris	UXO	mIoU
DeepLeng	0.164	0.913	0.060	0.763	0.475
Cuttlefish	0.115	0.744	-	0.150	0.336

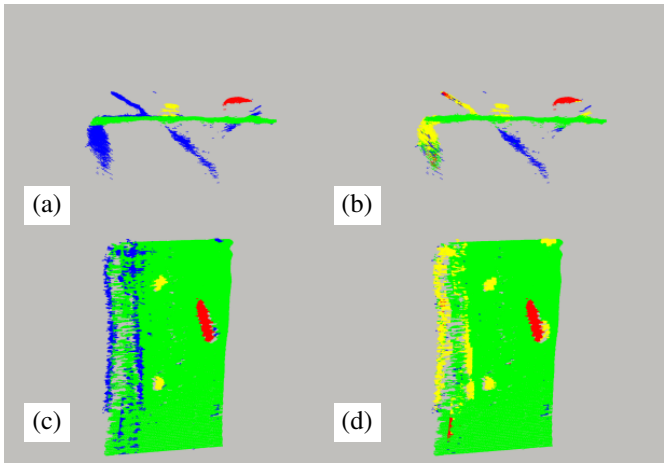

Fig. 7. Example segmentation result from the DeepLeng dataset. Manually annotated ground truth is shown in (a) and (c), which are respectively along-track and top-down views of the same data sample. The same across-track (b) and top-down (d) views are shown for the 3D segmentation result. Colored labels represent the classes of noise (blue), seabed (green), debris (yellow), and UXO (red).

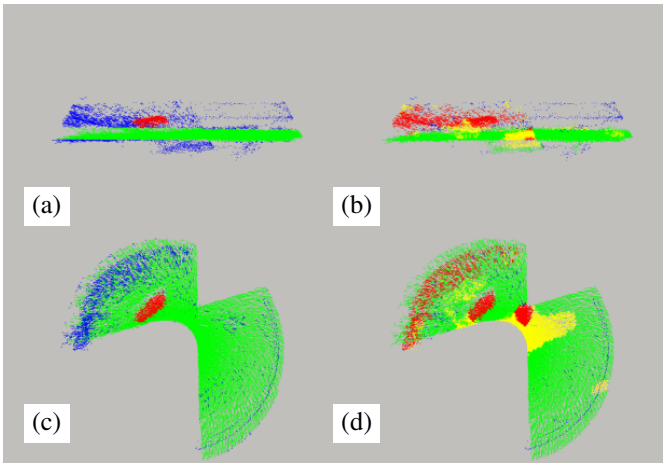

Fig. 8. Example segmentation result from the Cuttlefish dataset, where the Cuttlefish is performing an orbiting maneuver. Manually annotated ground truth is shown in (a) and (c), which are respectively along-track and top-down views of the same data sample. The same across-track (b) and top-down (d) views are shown for the 3D segmentation result. Colored labels represent the classes of noise (blue), seabed (green), debris (yellow), and UXO (red). Debris class is not present in the ground truth of this dataset, therefore any predicted points belonging to "debris" class are false positives.

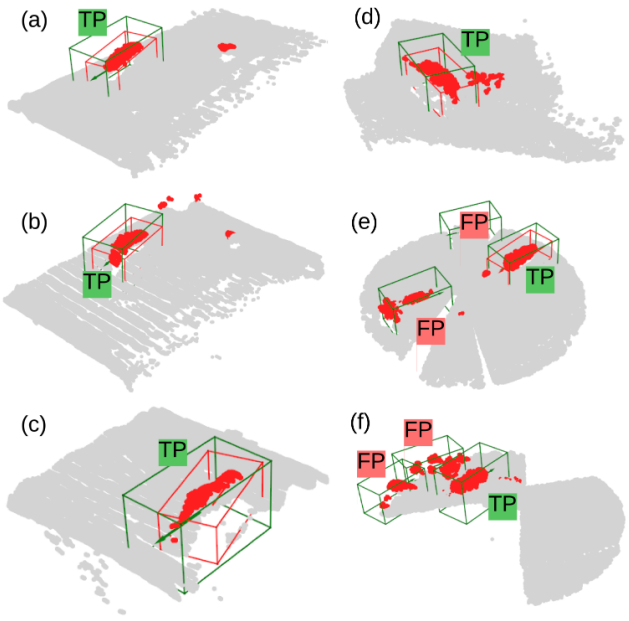

Fig. 9. 3D detection results from evaluation on real data. (a), (b), and (c) correspond to the DeepLeng dataset, while (d), (e), and (f) correspond to the Cuttlefish dataset. The green and red bounding boxes correspond to the model' prediction and the annotated groundtruth respectively. TP and FP cases are highlighted in each prediction result.

The model was trained for 80 epochs with a batch size of 1 on an NVIDIA GeForce RTX 4070 Ti GPU. For performance evaluation, the mean Average Precision (mAP) and recall metrics, which are commonly employed in object detection tasks, were used. The model achieved an mAP of 0.79 and a recall of 0.89 on the validation set, indicating strong detection performance.

Similar to the workflow used in the segmentation task, the 3D detection model was evaluated on the six real-world datasets described in IV-A. Because the model is designed to detect the target object rather than classify background points such as sea-floor structures or noise, an additional pre-processing step was applied. Specifically, statistical outlier filtering was performed to remove a significant portion of the noise points from the point clouds prior to inference.

Using the same evaluation metrics as in the simulation experiments, the 3D detection model achieved an average mAP of 0.69 at an IoU threshold of 0.5 across the six datasets. When evaluated on each platform dataset individually, the model achieved perfect detection performance (mAP = 1.0) on all three DeepLeng point clouds, accurately localizing the target UXO. In contrast, its performance on the Cuttlefish datasets was lower, with an average mAP of 0.375.

This performance drop can be attributed primarily to the distinct motion trajectories used during two of the Cuttlefish data acquisitions. These datasets employed an orbital scanning motion, which introduced higher noise levels and significantly altered the distribution and density of the captured points. This form of noise was not well represented in the synthetic training data, leading the model to produce several false positive predictions and, consequently, a reduction in mAP.

Nevertheless, it is important to note that despite this decrease in mAP, the model successfully detected the target UXO in all point clouds. Fig. 9 shows the results on all evaluated pointclouds. This result highlights the robustness of the detection model to previously unseen noise patterns and its ability to generalize to real-world data, albeit with a performance drop in more challenging scenarios. These findings suggest that incorporating noise profiles that better reflect real acquisition conditions into the training process may further improve the model's robustness and reduce false positives.

VI. Conclusion

This work has detailed how BLENDgänger can be used to generate synthetic MBES bathymetry datasets for the training of UXO perception models. Our experiments demonstrate that both 3D semantic segmentation and 3D object detection models can achieve satisfactory performance when trained on such synthetic data. However, the results also highlight the importance of accurately emulating real-world noise conditions within the simulation process. This becomes particularly evident with the Cuttlefish datasets, where orbiting maneuvers result in overlapping swaths and compounded noise artifacts. These maneuvers create local regions of high point density in the point cloud, a phenomenon not yet explicitly replicated in our simulation framework, which currently models more straightforward linear sensor trajectories. Nevertheless, BLENDgänger is sufficiently flexible to incorporate such complex maneuvers in future iterations, supporting more realistic simulation of challenging survey scenarios.

While this study has focused on models trained using purely bathymetric range measurements, incorporating backscatter intensity could further enhance perception model performance. Although BLAINDER supports backscatter simulation, realistic intensity values require accurately mapped material properties. With access to labeled real-world datasets containing backscatter information, BLAINDER's intensity simulation could be refined and augmented using empirical backscatter signatures, improving the fidelity and applicability of the synthetic data. This would be particularly useful for UXO detection, as the distinct metallic backscatter responses of UXO objects can help differentiate them from the surrounding seabed.

The future scope of BLENDgänger extends beyond MBES bathymetry simulation. Expanding the framework to include other data types, such as underwater imagery or additional sonar modalities would enable the development and evaluation of multi-modal perception models that fuse complementary sensor data. Such an approach could be especially beneficial for UXO detection, as the strengths of some sensors may compensate for weaknesses in others under varying environmental conditions, such as low light, high turbidity, or different burial depths. By facilitating multi-modal training, BLENDgänger has the potential to assist in the development of robust and generalizable perception systems for complex underwater environments.

The BLAINDER-based MBES simulation presented in this work, while effective for training purposes, can be further enhanced through the use of more physically realistic acoustic modeling. Owing to the modular design of BLENDgänger, generated scenes can be easily exported for use in external simulation engines such as ROS/Gazebo, Project DAVE, or Isaac Sim. This flexibility enables the adoption of advanced, high-fidelity acoustic models, supporting the creation of even more realistic synthetic datasets for perception model development.

The flexibility and extensibility of the BLENDgänger framework open avenues for further research and application beyond UXO detection. Potential future directions include adapting the platform for tasks such as underwater infrastructure inspection, habitat mapping, or other marine object detection challenges. As synthetic datasets become more representative of real-world complexities, the reliability, robustness and generalizability of machine learning models trained with such data will continue to improve, advancing autonomous perception capabilities in subsea environments. Ultimately, the advancement of synthetic dataset generation tools like BLENDgänger can help ensure that new machine learning-based detection methods are both reliable and ready for real-world deployment. This will not only enhance safety and reduce the risk of UXO-related incidents but also lead to significant cost savings in surveying and remediation, benefiting both governmental agencies and commercial operators working in marine environments.

REFERENCES

[1] A. J. Beck, M. Gledhill, C. Schlosser, B. Stamer, C. Böttcher, J. Sternheim, J. Greinert, and E. P. Achterberg, "Spread, behavior, and ecosystem consequences of conventional munitions compounds in coastal marine waters," *Frontiers in Marine Science*, vol. 5, no. APR, 2018.

[2] A. J. Beck, M. Gledhill, U. Gräwe, M. Kampmeier, A. Eggert, C. Schlosser, B. Stamer, J. Greinert, and E. P. Achterberg, "Widespread environmental contamination from relic munitions in the southwestern baltic sea," *Chemosphere*, vol. 372, p. 144115, 2025.

[3] M. Kampmeier, P. Michaelis, D. Wehner, T. Frey, M. Seidel, J. Wendt, and J. Greinert, "Workflow towards autonomous and semi-automized uxo survey and detection," *Proceedings of Meetings on Acoustics*, vol. 44, no. 1, p. 070025, 11 2021.

[4] J. Le Deunf, N. Debese, T. Schmitt, and R. Billot, "A review of data cleaning approaches in a hydrographic framework with a focus on bathymetric multibeam echosounder datasets," *Geosciences*, vol. 10, no. 7, p. 254, 2020.

[5] L. Arge, K. G. Larsen, T. Mølhave, and F. van Walderveen, "Cleaning massive sonar point clouds," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, pp. 152–161.

[6] A. S. et al., "Code for BLENDgänger: Generating Synthetic MBES Data for Underwater UXO Perception Tasks," 2025, available at: https://github.com/AmosSmith3/blendgaenger (accessed July 30, 2025). [Online]. Available: https://github.com/AmosSmith3/blendgaenger

[7] D. Henkel, E. González Ávalos, M. Kampmeier, P. Michaelis, and J. Greinert, "Machine learning as supporting method for UXO mapping and detection," EGU General Assembly 2020, 2020, presentation EGU2020-22594. [Online]. Available: https://doi.org/10.5194/egusphere-egu2020-22594

[8] C. Choy, J. Gwak, and S. Savarese, "4d spatio-temporal convnets: Minkowski convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[9] W. Wu, Z. Qi, and L. Fuxin, "Pointconv: Deep convolutional networks on 3d point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9621–9630.

[10] Y. Yan, Y. Mao, and B. Li, "Second: Sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, p. 3337, 2018.

[11] H. Wu, C. Wen, W. Li, X. Li, R. Yang, and C. Wang, "Transformation-equivariant 3d object detection for autonomous driving," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 3, 2023, pp. 2795–2802.

[12] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp. 77–85.

[13] K. Himri, P. Ridao, and N. Gracias, "3d object recognition based on point clouds in underwater environment with global descriptors: A survey," *Sensors*, vol. 19, no. 20, p. 4451, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/20/4451

[14] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," *IEEE International Conference on Robotics and Automation*, pp. 1–4, 2011.

[15] A. Aldoma, M. Vincze, N. Blodow, D. Gossow, S. Gedikli, R. B. Rusu, and G. Bradski, "Cad-model recognition and 6dof pose estimation using 3d cues," in *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Barcelona, Spain, 2011, pp. 585–592.

[16] M. Martin-Abadal, M. Piñar-Molina, A. Martorell-Torres, G. Oliver-Codina, and Y. Gonzalez-Cid, "Underwater pipe and valve 3d recognition using deep learning segmentation," *Journal of Marine Science and Engineering*, vol. 9, no. 1, p. 5, 2020.

[17] X.-X. Wang, J. Gao, and L. Feng, "Recognition and 3d pose estimation for underwater objects using deep convolutional neural network and point cloud registration," in *2020 International Conference on System Science and Engineering (ICSSE)*, Kagawa, Japan, 2020, pp. 1–6.

[18] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018. [Online]. Available: https://arxiv.org/abs/1804.02767

[19] Q. Hu, H. Zhu, M. Yu, Z. Fan, W. Zhang, X. Liu, and Z. Li, "A novel 3d detection system with target keypoint estimation for underwater pipelines," *Ocean Engineering*, vol. 309, p. 118319, 2024.

[Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801824016573

[20] M. Hinz, P. Westfeld, P. Feldens, A. Feldens, S. Themann, and S. Papenmeier, "Ai-based boulder detection in sonar data–bridging the gap from experimentation to application," *The International Hydrographic Review*, vol. 30, no. 1, 2024.

[21] N. Jaber, B. Wehbe, and F. Kirchner, "3d-duo: 3d detection of underwater objects in low-resolution multibeam echosounder maps," *Ocean Engineering*, vol. 331, p. 121254, 2025.

[22] J. P. Coffelt, A. Smith, N. Conen, and P. Kampmann, "Segmentation of multibeam echosounder bathymetry and backscatter," in *OCEANS 2024 - Singapore*, 2024, pp. 1–6.

[23] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, vol. 3, 2004, pp. 2149–2154 vol.3.

[24] L. Winiwarter, A. M. Esmorís Pena, H. Weiser, K. Anders, J. Martínez Sánchez, M. Searle, and B. Höfle, "Virtual laser scanning with helios++: A novel take on ray tracing-based simulation of topographic full-waveform 3d laser scanning," *Remote Sensing of Environment*, vol. 269, p. 112772, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425721004922

[25] NVIDIA, "Isaac Sim." [Online]. Available: https://github.com/isaac-sim/IsaacSim

[26] E. Potokar, K. Lay, K. Norman, D. Benham, T. Neilsen, M. Kaess, and J. Mangelson, "HoloOcean: Realistic sonar simulation," in *Proc. IEEE/RSJ Intl. Conf. Intelligent Robots and Systems, IROS*, Kyoto, Japan, Oct 2022.

[27] J. Song, H. Ma, O. Bagoren, A. V. Sethuraman, Y. Zhang, and K. A. Skinner, "Oceansim: A gpu-accelerated underwater robot perception simulation framework," 2025.

[28] W.-S. Choi, D. Olson, D. Davis, M. Zhang, A. Racson, B. Bingham, M. Mccarrin, C. Vogt, and J. Herman, "Physics-based modelling and simulation of multibeam echosounder perception for autonomous underwater manipulation," *Frontiers in Robotics and AI*, vol. 8, p. 706646, 09 2021.

[29] P. Cieślak, "Stonefish: An advanced open-source simulation tool designed for marine robotics, with a ros interface," in *OCEANS 2019 - Marseille*, jun 2019.

[30] M. Grimaldi, P. Cieslak, E. Ochoa, V. Bharti, H. Rajani, I. Carlucho, M. Koskinopoulou, Y. R. Petillot, and N. Gracias, "Stonefish: Supporting machine learning research in marine robotics," in *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, may 2025.

[31] S. Reitmann, L. Neumann, and B. Jung, "Blainder—a blender ai add-on for generation of semantically labeled depth-sensing data," *Sensors*, vol. 21, no. 6, 2021.

[32] M. Hildebrandt, S. Arnold, P. Kloss, B. Wehbe, and M. Zipper, "From epi- to bathypelagic: Transformation of a compact auv system for long-term deployments," in *2020 IEEE/OES Autonomous Underwater Vehicles Symposium (AUV)*, 2020, pp. 1–6.

[33] L. Christensen, J. Hilljegerdes, M. Zipper, A. Kolesnikov, B. Hülsen, C. E. S. Koch, M. Hildebrandt, and L. C. Danter, "The hydrobatic dual-arm intervention auv cuttlefish," in *OCEANS 2022, Hampton Roads*, 2022, pp. 1–8.