DOI: 10.23784/HN119-01

Automatic detection of boulders by neural networks

A comparison of multibeam echo sounder and side-scan sonar performance

An article by PETER FELDENS, PATRICK WESTFELD, JENNIFER VALERIUS, AGATA FELDENS and SVENJA PAPENMEIER

Neural networks show great promise in the automatic detection of boulders on the seafloor. Maps derived from bathymetric data show better performance compared to backscatter mosaics in this study. However, we find the lack of training data ground-truthed to a high standard the largest challenge for automated object detection based on acoustic data.

boulder detection | neural networks | hydrographic surveying | bathymetry | backscatter
Erkennung von Felsbrocken | neuronale Netze | Seevermessung | Bathymetrie | Backscatter

Neuronale Netze sind sehr vielversprechend bei der automatischen Erkennung von Felsbrocken auf dem Meeresboden. Aus bathymetrischen Daten abgeleitete Karten zeigen in dieser Studie eine bessere Leistung im Vergleich zu Rückstreumosaiken. Die größte Herausforderung für die automatische Objekterkennung auf Basis akustischer Daten ist jedoch der Mangel an Trainingsdaten, die auf einem hohen Standard erprobt sind.

Authors
Dr. Peter Feldens, Agata Feldens and Dr. Svenja Papenmeier work at the Leibniz Institute for Baltic Sea Research Warnemünde.
Dr. Patrick Westfeld and Jennifer Valerius work at the Federal Maritime and Hydrographic Agency (BSH) in Rostock and Hamburg.

peter.feldens@io-warnemuende.de

1 Introduction

Multibeam echo sounders (MBES) have been used for decades to provide high-quality bathymetric maps of the seafloor (Lurton 2002; Augustin et al. 1996; Pickrill and Todd 2003). The German Hydrographic Office (Federal Maritime and Hydrographic Agency, BSH) collects bathymetry and detects objects underwater by vessel-mounted MBES systems (Dehling and Ellmer 2012). The data surveyed in German waters are processed into official nautical charts and nautical publications to ensure navigational safety at sea. Accurate and reliable information of seabed's topography further forms a decisive basis for political and technical decisions relating to the sea, including applications depending on spatio-temporal-resolved 3-D geodata. Echo sounding is a measurement technique allowing for the 3-D reconstruction of the surface of the seafloor and all objects located on it. As a primary result, a digital surface model (DSM) is available. During the following data processing chain conducted at BSH, the task is to separate between the surface measured and the actual seabed, to derive a digital terrain model (DTM) of the seafloor. The detection and extraction of boulders are challenging. At BSH, it is realised in a semi-automatic process based on geometric filtering, with interactive post-processing and a final visual inspection by well-trained experts. This procedure is time-consuming and error-prone because of subjectivity and generalisation. Against a background of increasing user requirements (e.g. nautical information service needs a consistent separation of seabed and boulders for chart production; marine spatial planning requires information about conditions on the seabed to assess the impact of offshore construction projects) and the compliance with international standards (IHO S-44 Order 1a and 1b require the reliable detection of obstacles along all main shipping routes), automation of the processing chain is indispensable in terms of accuracy, reliability and reproducibility of the results. It is also required in the sense of an efficient evaluation of large areas.

Next to hydrographic applications, recent developments in habitat mapping require the detection of cobbles and larger hard substrates. The identification of marine hard substrates based on acoustic remote sensing is important for the detection, delineation and ecological assessment of seafloor habitats (Papenmeier et al. 2020) as well as for marine spatial planning. This need is accounted for in several international frameworks, such as the Convention on Biological Diversity and the Marine Strategy framework directives. Boulder detection in the German Baltic and the North Sea for these purposes is done using side-scan sonar (SSS) systems. Next to the ease of operation over large scales, the survey geometry of a side-scan sonar, towed above the seafloor,

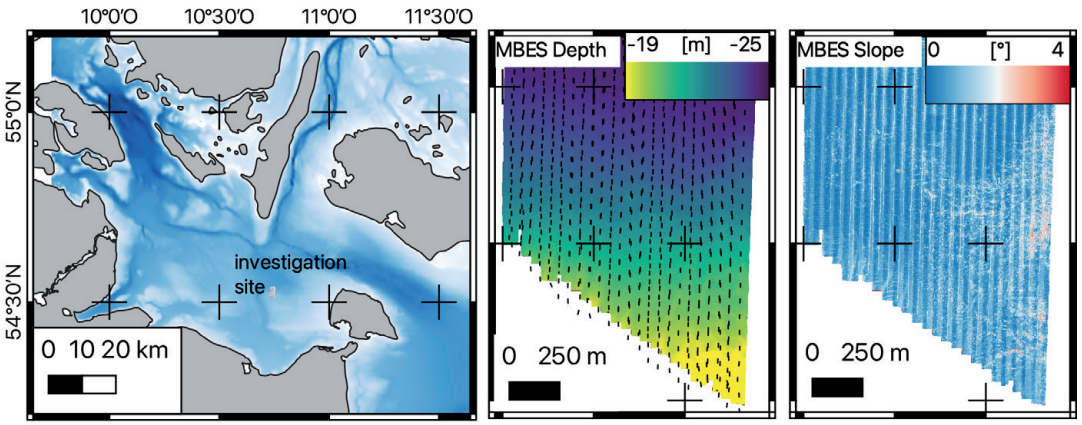

Fig. 1: Location of the investigation site west of Fehmarn (left). Water depths in the area range between 16 m and 25 m (centre), dashed lines are the survey lines run during MBES data collection. Right: Slope calculated from the local bathymetry

aids the detection of small objects. Both for manual and automatic methods, boulder detection was found to be more reliable, with an increasing number of pixels forming an object's representation in backscatter (BS) mosaics (Feldens et al. 2019). Acoustic shadows, which form behind boulders, increase the number of pixels of boulder representations in backscatter mosaics. Shadow sizes increase with grazing angle, thus favouring towed sonar systems (Papenmeier et al. 2020). Therefore, while the spatial resolution of modern MBES derived backscatter information can rival that of side-scan sonar systems in many relevant practical applications (depending on water depth), their survey geometry is unfavourable for boulder detection in backscatter data. However, the pixel-perfect co-registration of depth and backscatter and derived data sets may offset this disadvantage and facilitate boulder detection based on MBES data. Considering the interpretation of extensive areas, human experts have difficulties in combining information of multi-dimensional data sets, while machine learning algorithms are less limited by dimensionality and more efficient (Yokoya et al. 2017).

In the last decade, object detection frameworks based on convolutional neural networks (CNN) were applied to different topics, including remote sensing in the earth sciences (Ghamisi et al. 2017; Zhu et al. 2017) with great success. CNNs were used to find boulders in side-scan sonar backscatter mosaics, showing performance comparable to human experts in areas of moderate to good data quality (Feldens et al. 2019). It is the aim of this case study to compare the performance of multibeam echo-sounder and side-scan sonar to image boulders in single-band and multi-band data sets including depth, slope and backscatter intensity. An object detection framework based on a neural network is used to identify boulders in the data sets, and the results are compared with the interpretation of human experts.

2 Methods

2.1 MBES

Multibeam echo sounder data were collected in summer 2019 from the hydrographic survey vessel VWFS *Deneb,* operated by BSH, by a state-of-the-art MBES system Teledyne-Reson Seabat 7125-SV2. The system operates at 400 kHz with a 140° opening angle, a pulse length of 300 µs and 512 beams per swath. The seafloor of the study area (Fig. 1, left) was fully covered by 50 survey lines with 100 % overlap (Fig. 1, centre). The software Teledyne PDS was used for real-time data acquisition. A combined GNSS (Global Navigation Satellite Systems; good global but poor relative accuracy) and INS (Inertial Navigation System; good local accuracy but drifts without external reference) forms the basis for an accurate and reliable real-time direct georeferencing of MBES measurements. MBES instruments require an accurate portrayal of the sound speed structure of the water column. In this campaign, the distribution of water sound velocity was determined by continuous profile measurements using the multi-parameter online probe Sea & Sun Technology CTD 60Mc. Bathymetry data were processed using Teledyne CARIS HIPS & SIPS. The processing chain holds techniques for i.a. correction of sound velocity induced effects, calculation of a georeferenced 3-D point cloud, generation of 3-D surface representation of the bottom topography, outlier detection and filtering.

To create backscatter grids with a resolution of 0.25 m based on the multibeam echo sounder data provided as s7k-files, angular variations in intensities were removed using the open-source processing toolbox MB-System (Caress and Chayes 1996). A grazing angle of 40° (here, minor variations in incidence angle have little effect on backscatter intensity) was used as a reference angle. A low pass Gaussian mean filter stretching five samples in the across-track and three samples in the along-track direction was applied once to

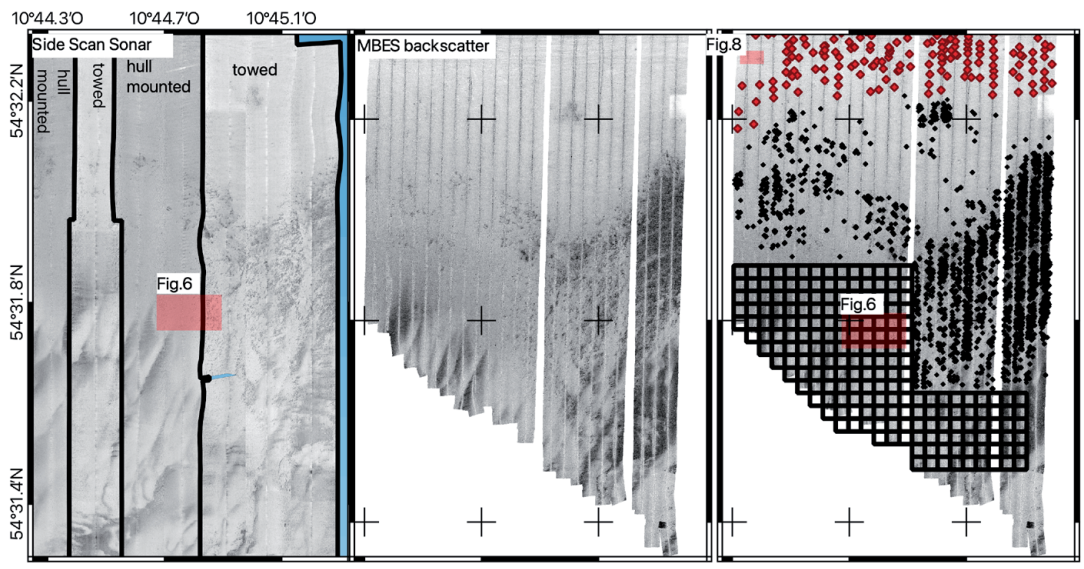

Fig. 2: Left: Side-scan sonar backscatter mosaic (0.25 cm pixel size) used for boulder classification in this study. Centre: Multibeam echo sounder backscatter mosaic. Right: Location of the boulders (black dots) and empty image examples (red dots) used for training of the models based on multibeam echo sounder data. The red box represents the test area for the manual identification of individual boulders in Fig. 6 and false positives examples in Fig. 8, respectively. The raster grid used for the manual determination of boulders densities is indicated

the data to remove high frequency speckle noise. Data gaps of up to 1.25 m were interpolated. The grid was built by applying a Gaussian Weighted Mean. As available profiles are overlapping, samples of higher grazing angles were given an increased priority during gridding. Profiles were run in both N-S and S-N directions on the same profile track. For the backscatter maps, one of these directions was used, the other line was discarded. Backscatter intensities were clipped at the 0.2 % and 99.8 % percentile to improve image contrast. In this study, higher backscatter intensities are displayed in darker colours. All backscatter intensities are uncalibrated, relative values (Lamarche and Lurton 2018) and were exported as 8-bit greyscale mosaics following processing. Multi-band images of MBES-derived grids of backscatter, slope and depth were created using the open source GDAL utilities (GDAL/OGR contributors 2021), by placing slope information in the green image channel (Fig. 1, right), backscatter information in the red image channel (Fig. 2, centre) and depth values in the blue image channel (Fig. 1, centre).

2.2 SSS

The side-scan sonar data were recorded in May 2019 during cruise #164 with the vessel VWFS *Deneb*. The Edgetech CSS-2000 was towed at an altitude of approximately 12 ± 3 m above the seabed. Due to technical problems with the CSS-2000 a change to the hull-mounted side-scan sonar (Edgetech 4300 MPX) became necessary during the cruise (Fig. 2 shows the coverage of both data sets). The vessel speed varied between 4 and 4.5 kn. Using a swath-width of 200 m the profile distance was set to 180 m to allow an overlap of approximately 10 % at the edges.

Processing of the backscatter amplitudes was done with the software package SonarWiz 7.3. Only the higher frequency of the CSS-2000 has been used (600 kHz). The 4300 MPX used a frequency of 410 kHz. After bottom tracking and empirical gain normalisation, the data of the towed system additionally required a correction of the navigation data. The sheave offset was adjusted, and a layback correction was executed basing on data of a cable counter and a pressure sensor. To generate a final backscatter mosaic both data sets were merged. The overlapping profiles were cut at the edges as far as possible without causing gaps. Finally, a mosaic (8-bit greyscale) with a spatial resolution of 25 cm was exported (Fig. 2, left).

2.3 Manual boulder count

Two experienced human interpreters did a manual count of individual boulders in a test area (Fig. 2, red box) based on the side-scan sonar mosaics. Human interpreters generally recognise boulders by an increased backscatter intensity facing towards the side-scan sonar, followed by an acoustic shadow forming behind. The human interpreters were not involved in picking the training data for the neural networks. To interpret larger areas, a raster approach is used. For 50 m × 50 m cells (Fig. 2, black raster grid), the same human experts decided whether it includes no boulders, one to five boulders, or over five boulders. This procedure is in line with currently published recommenda-

tions for mapping geogenic reefs (Heinicke et al., in press), used to characterise geogenic reefs over larger areas. The agreement between the human experts is calculated using the F_1 score of the resulting confusion matrix. An F_1 score of 1.0 indicates perfect agreement, while the lowest value is 0, when either precision or recall are 0. The F_1 score is calculated from the confusion matrix by $F_1 = 2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$. Values for each class (no boulders, one to five boulders and more than five boulders) were averaged.

2.4 Automatic boulder count

2.4.1 Neural network

Artifical neural networks are composed of series of interconnected layers of artificial neurons. In a trained neural network, input signals are transformed by changing weights at each connection, until the last layer of the network reports the result of the computation. Convolutional neural networks are a subset of neural networks and were developed for image classification with overwhelming success. While the architecture of CNNs varies, all include a series of convolutional layers, that operate by convolving a small part, often 3 × 3 pixels, of the underlying image (or the output of an earlier layer in the network) with weights initialised at random. This assumes that pixels in close vicinity are more likely to form patterns significant for the image context than those pixels with greater distance. The weights are adjusted during model training with annotated images to minimise a loss function. Loss functions compare the predictions of the neural network to the annotations. To allow CNNs learning non-linear features, activation functions change the output of layers in the network, while regular downsampling of the image size allows the network to learn features of larger scales.

The automated boulder count was done using the YOLO (You Look Only Once) framework, developed by Joseph Redmon (Redmon et al. 2015), with the current implementation available under a permissive license on GitHub (https://github.com/AlexeyAB/darknet). Lary et al. (2016) and Schmidhuber (2015) give a detailed description of convolutional neural networks and their application for image interpretation.

The YOLO network was developed for object detection. To identify and locate different objects in images is more complicated than the classification of entire images and requires a different network architecture. YOLO is a one-stage detector, meaning it analyses images in one pass (hence the abbreviation, You Only Look Once) while keeping high accuracy. One-stage detectors are a faster approach compared to other object detection frameworks that rely on multiple stages for object detection in images. The YOLO architecture is described by Bochkovskiy et al. (2020). In principle, it uses a series of different convolutional layers (the backbone and neck) to extract object features and divide the input image into grids at three different resolutions. For each grid cell at each resolution, it predicts the probability that the cell includes a learned object within anchor boxes of predefined size. These probabilities and the corresponding bounding box coordinates are the output of the trained model. YOLO networks are available in different configurations of the backbone, of which we here utilise the standard configuration of YOLO version 4.

2.4.2 Model training and application

To create the training data sets, a human interpreter identified bounding boxes of boulders in training areas in QGIS 3.16. Boulders were required to have a shadow. The boulders were exported as an SQLite database. The training database for the SSS model includes 13,847 boulder instances. A model was trained on a data set with an emphasis on small boulders comprising only a few pixels. This data set comprises 4,070 entries. The MBES training database was only started with the investigation site reported here (Fig. 2). It is not possible to use the same training data sets for MBES and SSS models, since the position accuracy of the side-scan sonar is not good enough to co-locate features of only a few pixels in size. Therefore, the MBES training data set comprises 2,654 instances of boulders (Fig. 2), with typical sizes of 3 × 3 to 8 × 15 pixels including shadows. The training mosaics were cut into small georeferenced images of 64 × 64 pixels (corresponding to approximately 16 m × 16 m in this study), overlapping by six pixels to minimise the number of training boulders that are cut by image boundaries. In the following, the pixel coordinates of the annotated examples were calculated and used as an input for training. Besides the annotated boulder examples, 182 examples of empty images (defined as containing no boulders) were selected for the MBES data set and 2,349 examples of empty images for the SSS data set.

For training, we used the YOLO network version 4, in contrast to earlier case studies that used the two-stage RetinaNet framework (Lin et al. 2017). We adhered to suggestions published on the project's GitHub page and changed the default configuration of the YOLO network. Therefore, the maximum number of training batches was reduced to 6,000 for MBES models and 24,000 for SSS models, the number of classes reduced to one, and the filter number of the convolutional layers before the object detection layers reduced to 18. Images were magnified to 512 × 512 pixels before training. Random variations in hue, exposure and saturation applied to the image were reduced from their standard settings to 0.1. The size of the input image was changed by 40 % every ten batches at random, and the size and aspect

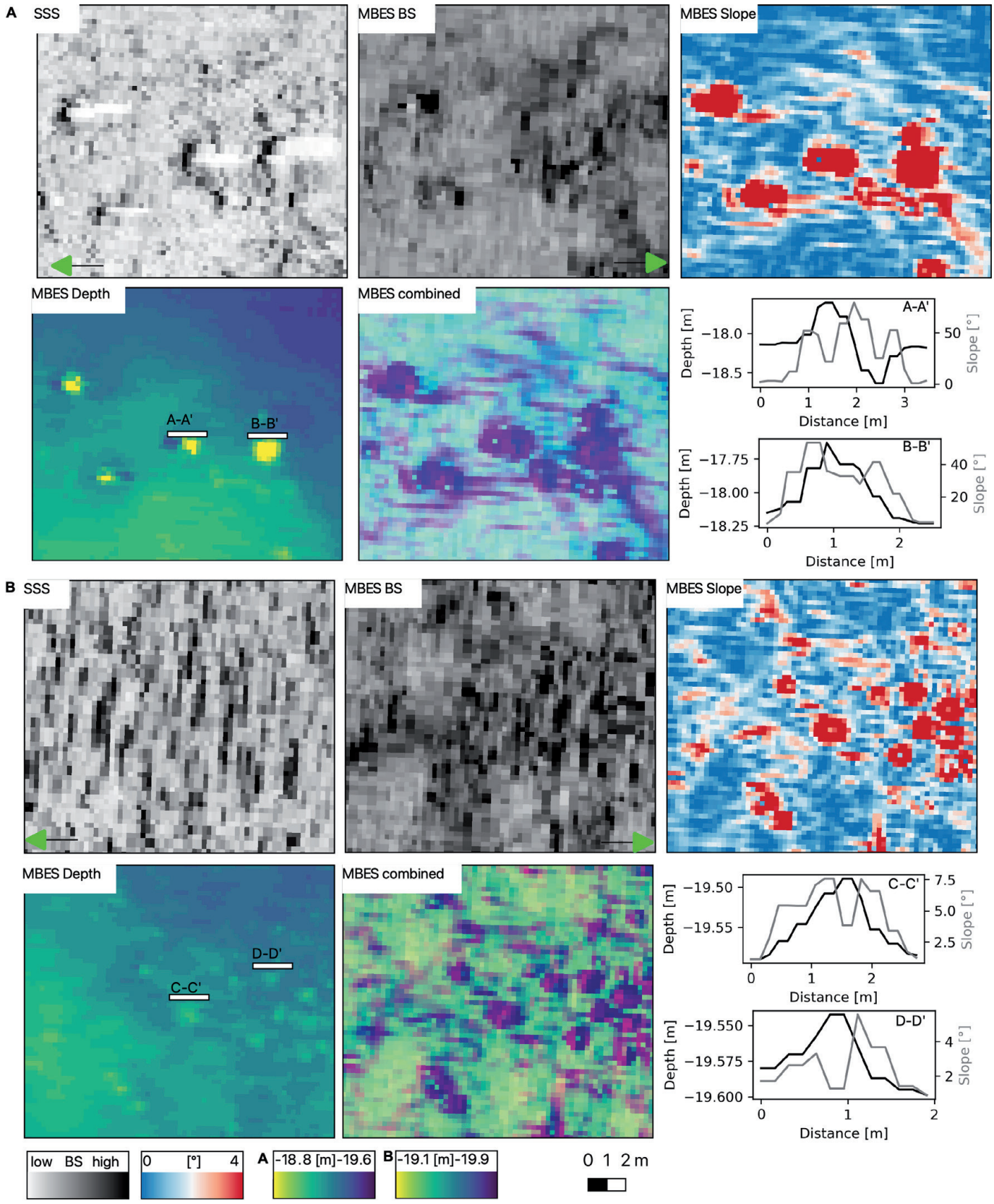

Fig. 3: The appearance of boulders in the different data sets. A) At a distance of 45 m to the nadir individual boulders are recognised in SSS backscatter. The same boulders (27 m to the nadir) are more difficult to recognise in MBES backscatter. The boulders are visible in bathymetry, slope, and combined data sets. B) Small boulders as imaged in the outer part (75 m to nadir) of a side-scan sonar swath. The characteristic boulder pattern is hard to recognise and appears smeared in the along-track direction, due to yaw movements or decreasing along-track resolution. The appearance of boulders is difficult to interpret in MBES (20 m to nadir) backscatter, but the objects are recognised in slope, bathymetry and combined data sets. The position of SSS images was shifted by several metres to account for positional differences to the MBES. The green arrow points to the nadir. SSS data was recorded with a CSS-2000

ratio were also changed by ±60 % for each image. The optimal anchor sizes for the YOLO network were calculated. 15 % of the training samples were randomly selected for validation and used to calculate the average precision for the boulder class (AP) of the different networks. After the image set for validation was separated, a Python script rotated every image in 45° steps to account for variable survey directions. The training took place on a NVIDIA 2080 TI graphic card (11 GB RAM). Training of the MBES models required about twelve hours for the MBES models and 40 hours for the large SSS model.

For model application, the training procedure is reversed. The (single or multi-band) mosaic is cut into small georeferenced image tiles of 64 × 64 pixels. Threshold values for include objects were set to 0.2 for all models except the SSS model for small objects, which was set to 0.35. The model is run on these small tiles. The detection of objects on a single image requires about 10 ms on an NVIDIA 2080 TI. The pixel-coordinates of the resulting bounding boxes are converted to geographic coordinates and displayed using QGIS. To emulate the raster approach used by human experts to cover large areas, detected boulders in each grid cell are counted.

3 Results

3.1 Local geology and appearance of boulders

Water depths in the investigation site (approximately 2 km^2) vary between 16 m and 25 m, with depths increasing towards the north. Backscatter maps derived from MBES and SSS show different seafloor facies at the site (Fig. 2), with fine-grained deposits and intensive disturbance by bottom trawling activities in the north (low backscatter). High backscatter intensities characterise glacial lag deposits towards the south and east. A high number of boulders are part of these deposits. Intermediate backscatter intensities towards south and west characterise fine to medium sands and partial outcrops of glacial lag deposits. In the side-scan sonar mosaics, which cover a larger area, a series of elongated, elevated ridges exist in the southeast. The general sedimentological build-up also controls the local slope shown in Fig. 1. While high pixel-to-pixel slopes exceeding 60° at maximum prevail in the areas of glacial lag deposits due to the presence of boulders and near the trawl marks, the remaining area is flat with slope values below 2°.

Based on a visual inspection, we find most boulders in the area composed of glacial lag deposits, with some also present in the sandy facies. The boulders have different characteristics in the data sets that are displayed in Fig. 3. In the SSS-derived backscatter mosaics, boulders can be recognised by a high backscatter front, an intermediate intensity signal behind and an acoustic shadow at the back, relative to the side-scan sonar position. However, small boulders are often more difficult to interpret. This is caused either by their small size or their position in the outer part of the swath (a combination of which is shown in Fig. 3B). In addition, artefacts in side-scan sonar data can resemble smaller boulders. Such artefacts include scatter from water column stratification or areas near the side-scan sonar nadir.

In MBES-derived backscatter, boulders are recognised by an increase in backscatter intensity compared to the surrounding seafloor (Fig. 3) but are often lacking a pronounced acoustic shadow. The backscatter representation of boulders is less distinct compared to SSS imagery in close to intermediate distance to the nadir. Boulders are imaged as circular to elliptic features in maps of the local slope. Slope values for boulders range from 3.5° to more than 60° degrees, related to the large variety of boulder shapes in transported lag deposits transported by glaciers. Also, boulders may be partially buried in the subsurface. However, not all circular features correspond to increased backscatter intensities, for example in the areas of overlapping profiles. In MBES-derived maps of depth, boulders are displayed as circular features elevated 2.5 cm to over 50 cm compared to the adjacent seafloor.

3.2 Manual boulder identification

For a test area of about 30,000 m^2, two experienced human interpreters picked boulders on the side-scan sonar backscatter mosaic (Fig. 4). The test area showcases instances of water column

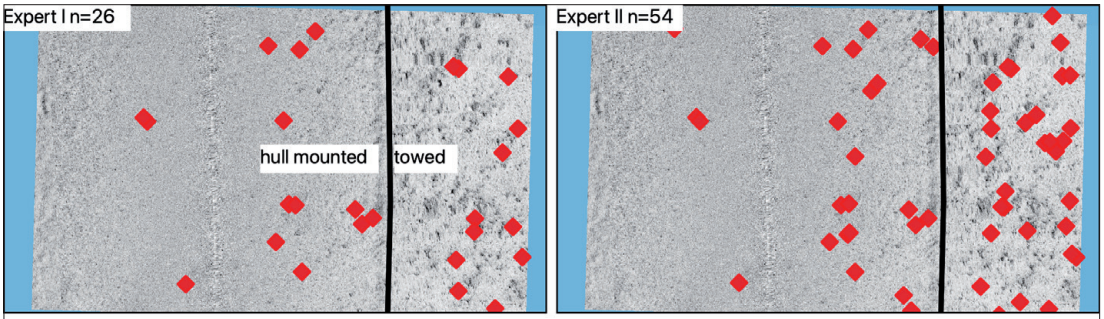

Fig. 4: Manual interpretation of boulder occurrence in the test area based on SSS backscatter data. The number of identified objects is 26 and 54. Refer to Fig. 2 for location

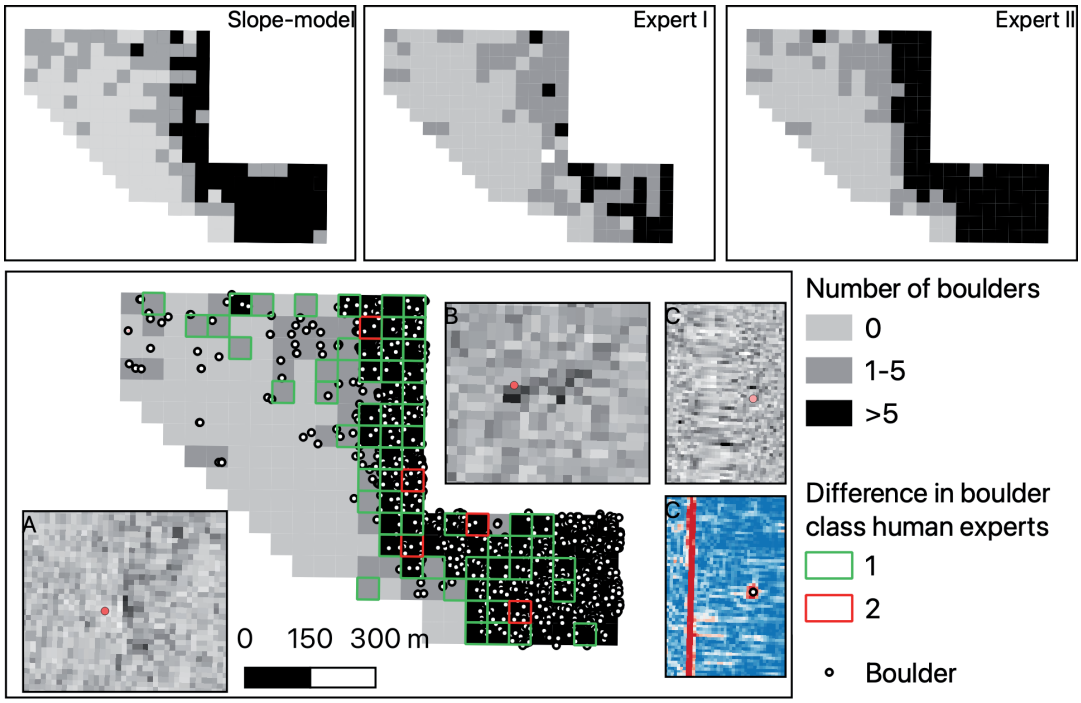

Fig. 5: Top: Number of boulders identified with the raster approach by the slope-model and to human experts. Bottom: Individual detection of the slope-model are plotted on top of the expert II interpretation. Coloured cell boundaries visualise the difference in interpretation between the human experts. A) Example of a potential boulder not noticed by the experts. B) A potential false positive detection of the model. C) Detections near the side-scan sonar nadir, where no judgment of the model detections is possible. For C, the slope map is shown in addition

stratification on the eastern side, a nadir stripe in the centre of the area and an overlap of two different profiles recorded with different side-scan sonars towards the west. The experts found 26 and 54 boulders. No human misinterpreted the water column artefacts, nadir stripes or overlapping profiles as boulders. A higher variability exists in the outer parts of the swath near the overlapping profiles, where the appearance of potential boulders varies. The same human experts interpreted boulder densities over a larger area using the raster approach applied to 50 m × 50 m cells (Fig. 5). Dense boulder assemblages were confirmed in the east towards the outcropping glacial till, while boulders are sparse towards west. Corresponding to the different number of individual boulders found in the test area, expert I identified a larger area covered by one to five boulders compared to expert II. The F_1 score, measuring the agreement between the two experts, is 0.61 based on 196 raster cells.

Data set	Model AP
MBES SLOPE	64 %
MBES DEPTH SLOPE BACKSCATTER	61 %
SSS BACKSCATTER large objects	43 %
SSS BACKSCATTER small objects	37 %
MBES DEPTH	36 %
MBES BACKSCATTER	18 %

Table 1: Overview of performance on the validation data set (measured in AP) for the different models and data sets

3.3 Automated boulder detection

The Average Precision (AP) of the models on the validation data is shown in Table 1. The highest performance is 64 % by the slope-only model, followed by a model working on a 3-band data set comprising MBES backscatter, slope and depth with 61 % AP. The MBES backscatter-only model achieves an AP of 18 %. The side-scan sonar performance is 37 % to 43 %, with the lower AP for the training data set with a focus on small objects. The detections of the best-performing slope-model are plotted on top of boulder densities as determined by human experts (Fig. 5).

The resulting detections of the models in the test area are shown in Fig. 6. The SSS models find a total of 35 boulders, all including a discernible shadow on visual inspection. One likely false positive occurs around water column stratification artefacts and one false positive in the nadir region. The MBES backscatter model finds a total of 29 boulders. Of these, seven have no discernible shadow, while the remaining display at least one pixel of acoustic shadows behind. The model working on the area-wide bathymetric grids detects 14 boulders with elevations of 6 cm to 40 cm compared to the surrounding seafloor, albeit most boulders smaller than 15 cm are not recognised in the data set. The slope model finds

59 objects at the test site, characterised by slopes ranging from 35° to less than 3.5°. However, most identified boulders show slope values of over 4°. The model running on the combined data set of backscatter, slope and depth detects 53 boulders. Most of these boulders are also recognised in the slope data set. However, several potential boulders found in the slope data set were not found by the combined model and vice versa, with examples shown in Fig. 7. Here, a comparison with the independently recorded side-scan sonar data – barring some uncertainty because of the positional

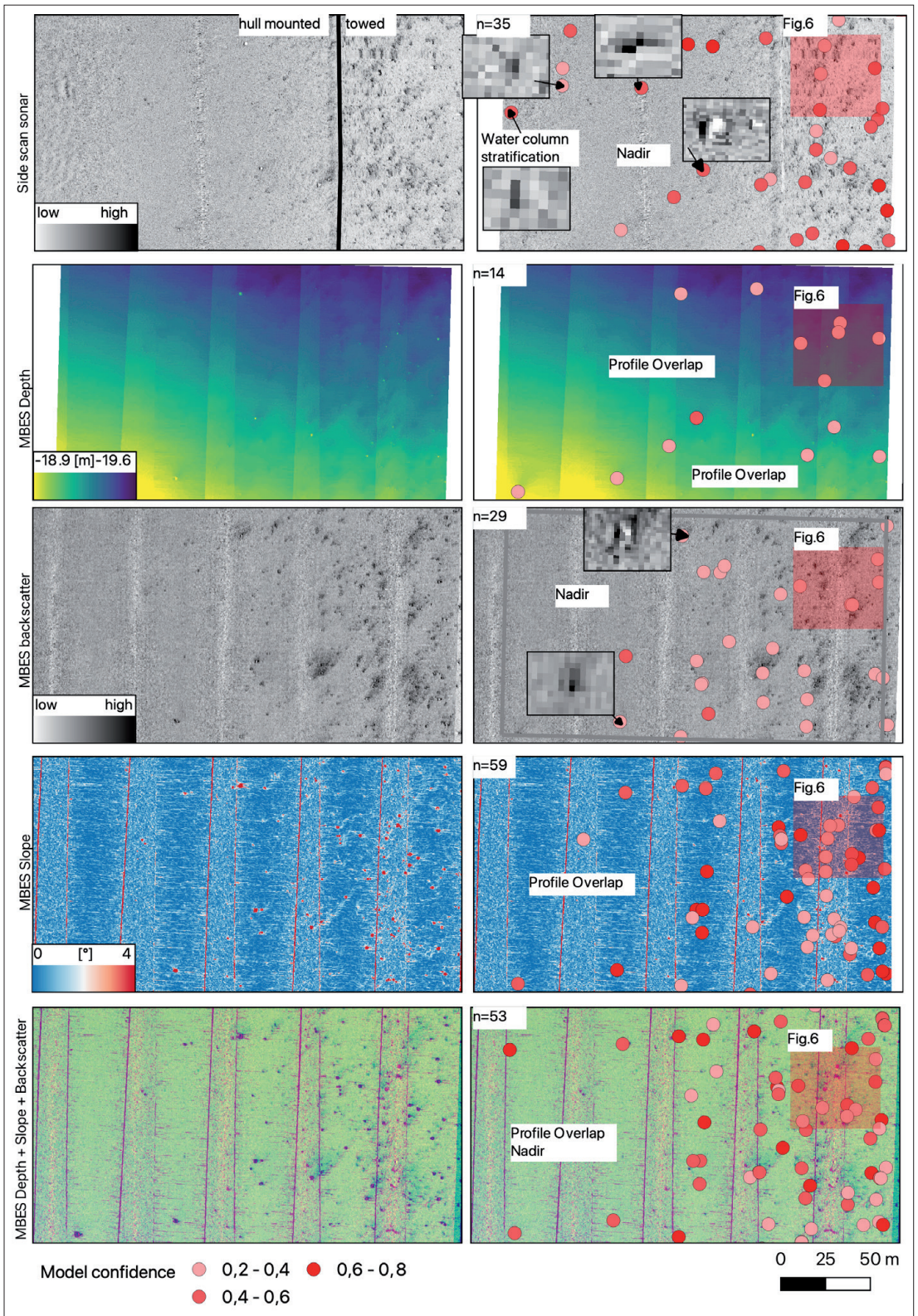

Fig. 6: Boulders found by the models in the test area in the different data sets. For the SSS backscatter mosaic, magnified insets show the similarity of small boulders and artefacts due to water column stratification and near-vertical incidence. Refer to Fig. 2 for location

inaccuracy that required shifting the side-scan sonar mosaic location by a few metres – seems to show that the slope data is correct, and these objects should have been identified as boulders. In contrast, in the northern test area (Fig. 8), circular elevated features are identified as boulders by the slope-model. We find similar examples, not displayed here, in areas with remaining outliers in morphological data which have a similar appearance. Such outliers cause artificial slopes but do not affect backscatter data information.

The results of the raster approach using the model with the highest AP (the slope-model) are shown in Fig. 5. The slope-model identifies between 0 and 42 boulders in the 50 × 50 m cells. The agreement with the human experts I and II as measured by the F_1 score for 182 cells (cells where both SSS and MBES data are available) is 0.75 and 0.63, respectively.

4 Discussion

The high difference of boulder detection by very experienced human interpreters (Fig. 3) shows the need for an objective, automatic method for boulder detection. The different count of individual boulders transfers to an agreement of 0.61 (F_1 score) over 196 cells that were interpreted with the raster approach. This poses a significant challenge both for quantification of model performance and for the establishment of correctly annotated training images, a problem faced by many other applications of neural networks to remote sensing data (Zhu et al. 2017). The same person interpreting the training database and the reference sites for boulder detection (Feldens et al. 2019) partially mitigates the problem. However, this approach does not scale to more than one involved person or to applications where objective results without interpreter bias are required. Almost no study includes an extensive ground truthing for boulders

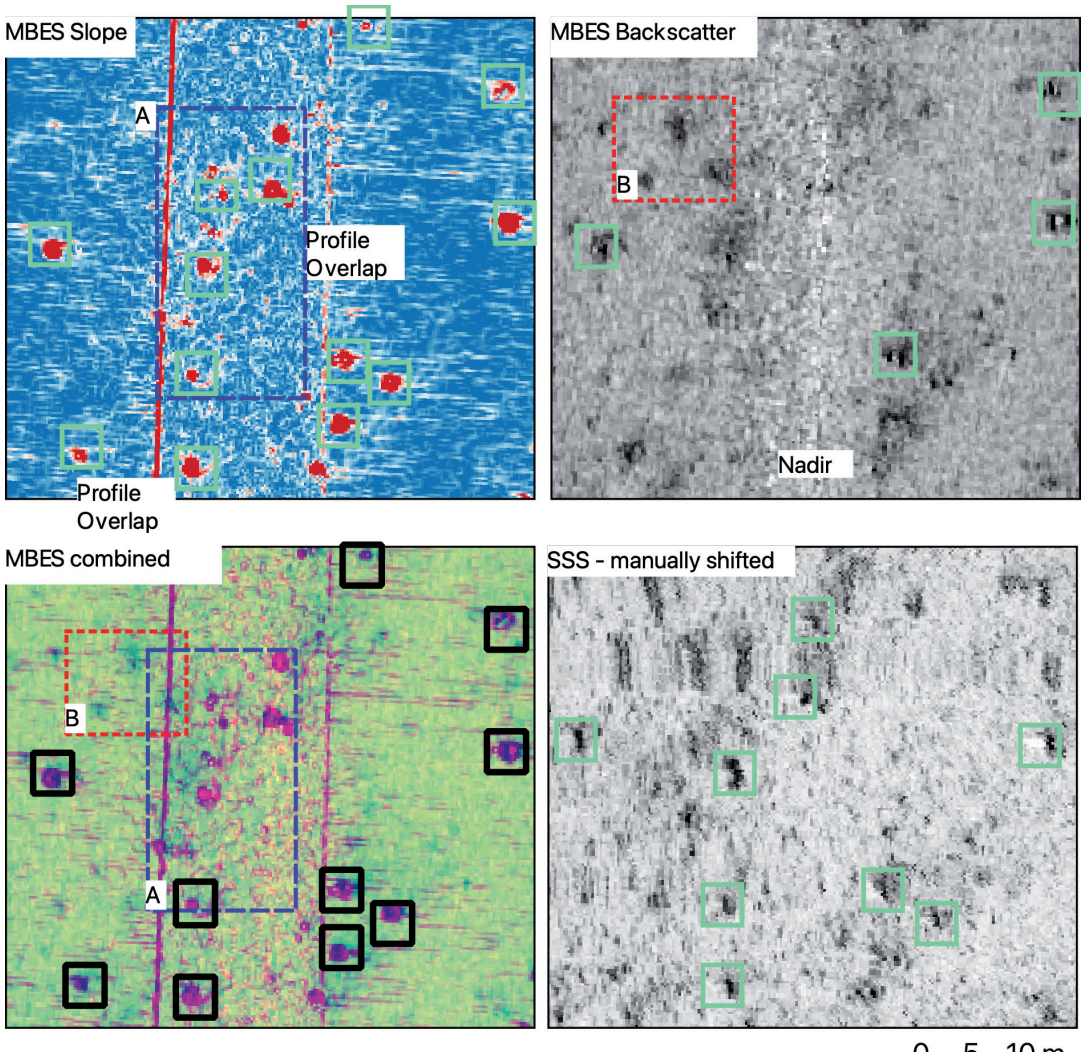

Fig. 7: Boulders detected by the MBES-models are displayed. Boulders are verified in the side-scan sonar image, whose position was shifted to account for positional inaccuracies. Near the nadir, potential boulders are not imaged in MBES backscatter data, while present in the slope map (blue rectangle). Vice versa, the backscatter map displays increased backscatter intensities in areas where no increased slope exists (red rectangle). No boulders are detected in both areas by the combined model working on depth-slope and backscatter channels. Refer to Fig. 6 for colour scales

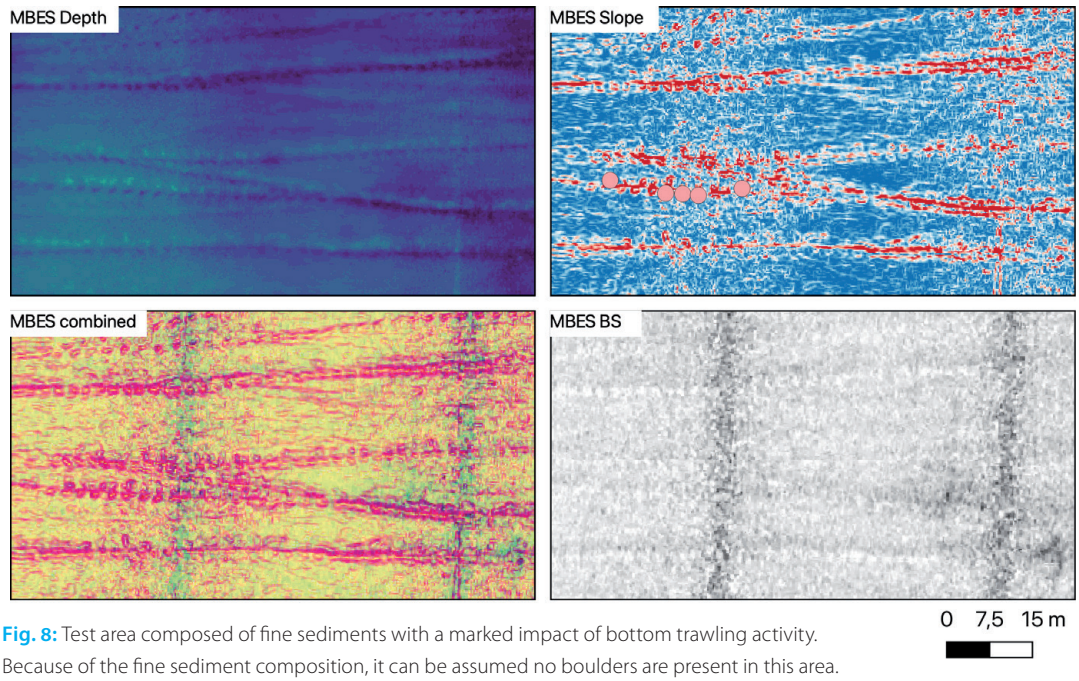

Fig. 8: Test area composed of fine sediments with a marked impact of bottom trawling activity. Because of the fine sediment composition, it can be assumed no boulders are present in this area. The model working on slope data detects several false positives in the area, while models running on backscatter, depth and the combined multi-band image report no false positives. Refer to the northwest of Fig. 2 for location

in acoustic data, and – except for obvious instances – the interpretation of a human interpreter of what is and what is not a boulder varies based on his/her experience, with no possibility to judge what is the correct interpretation. The appearance and visibility of boulders in backscatter data can change with swath width and incidence angle (Papenmeier et al. 2020; von Rönn et al. 2019). While a methodological description on how to assess geogenic reefs exists (Heinicke et al., in press), it defines no sufficient criteria to decide which objects are to be identified as boulders in acoustic data.

Still, our case study allows qualitative insight into the advantages and disadvantages of SSS and MBES-based boulder mapping by neural networks. To mitigate the impact on AP for the different models, a single person confirmed all samples in the training database used for this study. Therefore, model performance is only compared relative to the interpretation of the acoustic data by one human expert and not to the true seafloor conditions. Both SSS and MBES systems supply backscatter information. A problem of SSS-based boulder detection are artefacts (Wilken et al. 2012), e.g., near the nadir or in areas of water column stratification that can in their structure resemble small boulders (Fig. 6). Due to the requirements to detect tiny objects comprising only 7 to 9 pixels in the examples shown here and even less if objects of 25 cm in size are to be detected in acoustic data (von Rönn et al. 2019), there is limited information to differentiate between artefacts and real objects. This causes a trade-off during the training of side-scan sonar-based models: if the sensitivity of the model to detect small boulders – as required by regulations – is increased, the amount of false positive identifications increases as well. Because of the absence of well ground-truthed reference sites, a calculation of meaningful precision-recall curves to find optimal threshold values is not possible. Tuning the threshold level of the model to the local conditions (e.g., the number of artefacts in the data) is done manually, which is a subjective procedure. A possible solution is to include nadir and water column stratification effects as distinct classes and define these areas as insufficient for boulder detection. While MBES snippet-derived backscatter data is not affected by water column stratification and is used for object detection (e.g., Kunde et al. 2018), individual boulders are not displayed in the specular regime (Fig. 7) at near-vertical incidence angle and are resolved in less detail compared to side-scan sonar images in the data (Fig. 2). The loss of detail may be caused by a different along-track resolution due to different opening angles of the used systems (0.5° at 400 kHz for the Reson 7125, CSS-2000: 0.26° at 600 kHz, respectively 0.29° at 410 kHz for the 4300 MPX). Combined with the less pronounced acoustic shadows, the AP of the MBES backscatter model data set, therefore, is worse compared to the model trained on side-scan sonar backscatter data (Table 1). MBES-based backscatter maps cannot be recommended as the principal data source for boulder detection based on our case study.

An obvious problem related to the use of MBES bathymetry and derived slope values is the required thorough cleaning of the data, with outli-

ers or morphological features having similarities to small boulders in slope maps. An example of such morphological feature in the German Baltic Sea is related to bottom trawling (Fig. 8). The trawl doors create steep local, almost circular morphological features when lifted off the seafloor. These features are misinterpreted by the slope-only model as boulders. The backscatter model correctly ignores these features. The combination of backscatter and slope data also prohibit false positives in the combined model. Therefore, while the AP of the slope model is the best on the validation data overall (Table 1), it also produced undesirable false positives in areas where boulders are very unlikely to appear (Fig. 7). The pixel-perfect coregistration of depth and backscatter information by multibeam echo sounders can mitigate this downside. Being the best model in our case study, the slope-model results were compared with the human raster-based interpretation of a larger area. The F_1 score of the model compared to the human experts is 0.75 and 0.62. Both scores are higher than the score for the direct comparison of the human experts, although the number of raster cells counted is not identical due to the different extension of available SSS and MBES data. Positional inaccuracies between the side-scan sonar and multibeam echo sounder data of approximately 5 m may negatively impact the comparison of cells where boulders are situated close to the edges. In hindsight interpretation of the model-human differences, potential errors on both sides were identified (examples shown in Fig. 5). In addition, the slope data is less affected than backscatter intensity by survey geometry and finds boulders that could not be identified in the side-scan sonar data because they are located close to the nadir.

The poorer performance of the MBES depth-derived model compared to the slope model is not surprising, given that the maximum resolution of the input image is the regional depth interval divided by the available discrete pixel values. In our study, this is 9 m divided by 256 (28 bpp, bit per pixel), artificially limiting the vertical resolution to ca. 0.035 m in the single band 8-bit image. Given that many boulders have smaller elevations (Fig. 2) and are visible in slope maps, the performance of the depth model is good and may have great potential for models operating on point clouds and derived statistics which became available in the last years (Held and Schneider von Deimling 2019; Guo et al. 2020). The advantages and disadvantages of including absolute depths as an input channel for neural networks must be considered, however. In the Baltic Sea, for example, finding boulders in deeper muddy basins is unlikely, but possible (Beisiegel et al. 2019), and changing depth intervals between different sites (and thus changing resolution in colour-coded depth images) may be problematic. We suggest exploring the use of further depth-derived information, such as the bathymetric position index, or texture parameters derived from backscatter mosaics in the future.

Models working on a combination of depth, slope and backscatter data produced false negatives in the near-nadir region, as boulders are not imaged in the backscatter channel. They also show fewer false positives and are less susceptible to remaining outliers in bathymetric data. Therefore, while the performance of the joint depth-slope-backscatter data set is worse than for the slope-only model (due to validation examples in the nadir region) in our case study, its inherent robustness to false positives by combining independent data sets makes it the method of choice for practical applications in the future. Ideally, and needed for many commercial applications anyway, an over 100 % overlap would remove the near-vertical incidence backscatter data and is expected to improve model results. Multi-band images with calibrated backscatter data collection (Lamarche and Lurton 2018) would also allow for a quantitative definition of boulders, e.g., by measuring increase of backscatter intensity in addition to local slope and local bathymetric position index.

5 Conclusion

Our case study shows that boulders are detected with higher precision in bathymetric data compared to backscatter mosaics recorded by either multibeam echo sounder or side-scan sonar. The results of the best model are comparable to the range of results achieved by human interpreters. We recommend combining bathymetry and backscatter data into a multi-band image to limit false positive detections. However, the limiting factor for the automated detection of boulders in acoustic data is not the technology, but the domain knowledge and the availability of accurately annotated training images. Future activities should involve the careful choice of sites for ground-truthing and acoustic surveys, to create a high-quality training data set. //

Acknowledgment

The authors thank Elham Al-Akrami for initial preparation of MBES-related training data set, and Merle Hennig for support in digitising boulders on the SSS mosaics. We thank the crew of VWFS *Deneb* for their great support during the measurement campaigns, and the two reviewers who provided helpful and constructive comments.

References

Augustin, Jean-Marie; Raymond Le Suavé et al. (1996): Contribution of the multibeam acoustic imagery to the exploration of the sea-bottom. Marine Geophysical Researches, DOI: 10.1007/BF00286090

Beisiegel, Kolja; Franz Tauber et al. (2019): The potential exceptional role of a small Baltic boulder reef as a solitary habitat in a sea of mud. Aquatic Conservation: Marine and Freshwater Ecosystems, DOI: 10.1002/aqc.2994

Bochkovskiy, Alexey; Chien-Yao Wang, Hong-Yuan Mark Liao (2020): YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv: 2004.10934

Caress, David W.; Dale N. Chayes (1996): Improved processing of Hydrosweep DS multibeam data on the R/V Maurice Ewing. Marine Geophysical Researches, DOI: 10.1007/BF00313878

Dehling, Thomas; Wilfried Ellmer (2012): Zwanzig Jahre Seevermessung seit der Wiedervereinigung. AVN Vol. 119, Nr. 7, S. 243–248

Feldens, Peter, Alexander Darr et al. (2019): Detection of Boulders in Side Scan Sonar Mosaics by a Neural Network. Geosciences, DOI: 10.3390/geosciences9040159

GDAL OGR contributors (2021): GDAL/OGR Geospatial Data Abstraction software Library. Open Source Geospatial Foundation

Ghamisi, Pedram; Javier Plaza et al. (2017): Advanced Spectral Classifiers for Hyperspectral Images: A review. IEEE Geoscience and Remote Sensing Magazine, DOI: 10.1109/MGRS.2016.2616418

Guo, Yulan; Hanyun Wang et al. (2020): Deep learning for 3d point clouds: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, DOI: 10.1109/TPAMI.2020.3005434

Heinicke, Kathrin; Tim Bildstein; Dieter Boedecker (in press): Leitfaden zur großflächigen Abgrenzung und Kartierung des LRT 1170 »Riffe« in der deutschen Ostsee (Untertyp: geogene Riffe)

Held, Philipp; Jens Schneider von Deimling (2019): New Feature Classes for Acoustic Habitat Mapping – A Multibeam Echosounder Point Cloud Analysis for Mapping Submerged Aquatic Vegetation (SAV). Geosciences, DOI: 10.3390/geosciences9050235

Kunde, Tina; Philipp Held et al. (2018): Ammunition detection using high frequency multibeam snippet backscatter information. Marine Pollution Bulletin, DOI: 10.1016/j.marpolbul.2018.05.063

Lamarche, Geoffray; Xavier Lurton (2018): Recommendations for improved and coherent acquisition and processing of backscatter data from seafloor-mapping sonars. Marine Geophysical Research, DOI: 10.1007/s11001-017-9315-6

Lary, David J.; Amir H. Alavi et al. (2016): Machine learning in geosciences and remote sensing. Geoscience Frontiers, DOI: 10.1016/j.gsf.2015.07.003

Lin, Tsung-Yi; Priya Goyal et al. (2017): Focal Loss for Dense Object Detection. arxiv: 1708.02002

Lurton, Xavier (2002): An introduction to underwater acoustics: principles and applications. Springer Science & Business Media

Papenmeier, Svenja; Alexander Darr et al. (2020): Hydroacoustic Mapping of Geogenic Hard Substrates: Challenges and Review of German Approaches. Geosciences, DOI: 10.3390/geosciences10030100

Pickrill, Richard A.; Brian J. Todd (2003): The multiple roles of acoustic mapping in integrated ocean management, Canadian Atlantic continental margin. Ocean & Coastal Management, DOI: 10.1016/S0964-5691(03)00037-1

Redmon, Joseph; Santosh Divvala et al. (2015): You Only Look Once: Unified, Real-Time Object Detection. DOI: 10.1109/CVPR.2016.91

Schmidhuber, Juergen (2015): Deep learning in neural networks: An overview. Neural Networks, DOI: 10.1016/j.neunet.2014.09.003

von Rönn, Gitta Ann; Klaus Schwarzer et al. (2019): Limitations of Boulder Detection in Shallow Water Habitats Using High-Resolution Sidescan Sonar Images. Geosciences; DOI: 10.3390/geosciences9090390

Wilken, Dennis; Peter Feldens et al. (2012): Application of 2D Fourier filtering for elimination of stripe noise in side-scan sonar mosaics. Geo-Marine Letters, DOI: 10.1007/s00367-012-0293-z

Yokoya, Naoto; Claas Grohnfeldt; Jocelyn Chanussot (2017): Hyperspectral and Multispectral Data Fusion: A comparative review of the recent literature. IEEE Geoscience and Remote Sensing Magazine, DOI: 10.1109/MGRS.2016.2637824

Zhu, Xiao Xiang; Devis Tuia et al. (2017): Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, DOI: 10.1109/MGRS.2017.2762307