Oceanologia 66 (2024) 286—298

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.journals.elsevier.com/oceanologia

ORIGINAL RESEARCH ARTICLE

Coverage estimation of benthic habitat features by semantic segmentation of underwater imagery from South-eastern Baltic reefs using deep learning models

Andrius Šiaulys[a,*], Evaldas Vaičiukynas[b], Saulė Medelytė[a], Kazimieras Buškus[b]

[a] *Marine Research Institute, Klaipeda University, Klaipeda, Lithuania*
[b] *Faculty of Informatics, Kaunas University of Technology, Kaunas, Lithuania*

Received 10 July 2022; accepted 17 December 2023
Available online 4 January 2024

KEYWORDS
Underwater video;
Epibenthos;
Mosaicking;
Segmentation;
Computer vision,
PSPNet;
ResNet;
Baltic Sea

Abstract Underwater imagery (UI) is an important and sometimes the only tool for mapping hard-bottom habitats. With the development of new camera systems, from hand-held or simple "drop-down" cameras to ROV/AUV-mounted video systems, video data collection has increased considerably. However, the processing and analysing of vast amounts of imagery can become very labour-intensive, thus making it ineffective both time-wise and financially. This task could be simplified if the processes or their intermediate steps could be done automatically. Luckily, the rise of AI applications for automatic image analysis tasks in the last decade has empowered researchers with robust and effective tools. In this study, two ways to make UI analysis more efficient were tested with eight dominant visual features of the Southeastern Baltic reefs: 1) the simplification of video processing and expert annotation efforts by skipping the video mosaicking step and reducing the number of frames analysed; 2) the application of semantic segmentation of UI using deep learning models. The results showed that the annotation of individual frames provides similar results compared to 2D mosaics; moreover, the reduction of frames by 2—3 times resulted in only minor differences from the baseline. Semantic segmen-

* Corresponding author at: Marine Research Institute, Klaipeda University, Klaipeda, Lithuania.
E-mail address: andrius.siaulys@ku.lt (A. Šiaulys).

Peer review under the responsibility of the Institute of Oceanology of the Polish Academy of Sciences.

Production and hosting by Elsevier

https://doi.org/10.1016/j.oceano.2023.12.004

tation using the PSPNet model as the deep learning architecture was extensively evaluated, applying three variants of validation. The accuracy of segmentation, as measured by the intersection-over-union, was mediocre; however, estimates of visual coverage percentages were fair: the difference between the expert annotations and model-predicted segmentation was less than 6—8%, which could be considered an encouraging result.

1. Introduction

Renewable energy installations, oil and gas drilling, maritime shipping and fishing, ecosystem surveillance and biodiversity conservation, aquaculture production, and a variety of other uses are becoming more common and increasing the need for maritime space. The need for maritime space necessitates integrated planning and management strategies based on sound scientific understanding and accurate seabed mapping (Smith and Cardoso, 2020), with underwater images (Urra et al., 2021) being one of the most widely used seabed mapping materials. The main advantage of underwater imagery is its cost-effectiveness and simplicity, which allow for the rapid collection of large volumes of data with a variety of underwater cameras, from relatively simple handheld GoPros or “drop-down” cameras to more advanced ROV and AUV-mounted filming systems. There are several applications and platforms designed or utilized for underwater imagery analysis, such as BIIGLE 2.0 (Langenkämper et al., 2017), CPCe (Kohler and Gill, 2006), Image J (Ferreira and Rasband, 2012), PhotoQuad (Trygonis and Sini, 2012), and broad-scale projects ongoing collecting huge amounts of video material, e.g., MAREANO (Buhl-Mortensen et al., 2015), yet only a small part of the information is being extracted due to labour-intensive and time-consuming analysis procedures. A promising way to process large amounts of images is computer-aided analysis, i.e., conversion of raw seabed video to 2D mosaics (Casoli et al., 2021; Šaškov et al., 2015), annotation and image segmentation (Martin-Abadal et al., 2018; Piechaud et al., 2019; Šiaulys et al., 2021), and quantification of segmentation results (Buškus et al., 2021).

Automatic segmentation of underwater imagery, compared to other types of image analysis, is a relatively new and challenging research direction. According to a survey (Gracias et al., 2017), the first publications on the seabed segmentation task (also termed seafloor classification) appeared 25 years ago and are still scarce, the common ground between them being the use of “hand-crafted” image features and traditional machine learning algorithms, for example, random forest (Rimavičius et al., 2018). Novel deep learning architectures of neural networks could be the enabling technologies to replace image features and analyse images more effectively, accurately, and quickly than ever before. Initial efforts to apply deep learning to UI concern corals (Alonso et al., 2019) and other broad categories such as fish, plants, divers, or stones (Islam et al., 2020; Liu and Fang, 2020), mostly from independent photographs and with little preoccupation with the sea floor. Other studies have shown that seafloor videos could be converted into 2D mosaics (multiple frames that are stitched together into a single still image), which can later be used for efficient visual analysis (Medelytė et al., 2022a; Šaškov et al., 2015) with applications of deep learning models (Buškus et al., 2021). However, the mosaicking of seabed videos is a labour-intensive process, requiring specific software and professional knowledge (Li et al., 2019), with some prerequisites for video material as well (stable distance from the seabed, homogeneous lighting, no fast-moving objects, etc.), which can be unachievable in rough open seas or very dynamic coastal areas.

In this study, two ways of making UI analysis more efficient were tested on eight dominant visual features of the SE Baltic reefs: 1) the simplification of video processing and expert annotation effort by skipping the video mosaicking step and reducing the frames analysed, and 2) the application of semantic segmentation of UI using deep learning models for automatic estimates of seabed visual features. Estimations were done both manually, by obtaining expert annotations, and automatically, by training a deep learning convolutional architecture on the annotated data. Experiments measure segmentation success and accuracy of automated visual coverage estimates through three types of validation: two-fold cross-validation, leave one out validation, and hold-out validation.

2. Material and methods

2.1. Underwater imagery data

Underwater videos were filmed in the coastal and offshore reefs of Lithuanian marine waters in the South-Eastern Baltic Sea at eight locations (Figure 1). Underwater video filming was carried out 1 m above the seabed, at depths of 4—8 m in the coastal area and 30—40 m offshore. The underwater videos in the coastal area were collected by SCUBA divers with a handheld GoPro underwater camera and “drop-down” type camera system equipped with an analog camera with 700 TV lines (TVL) resolution for live view and a digital camera (Panasonic HX-A500) that recorded the seabed at high resolution (1280 × 720 px). Offshore data was collected using an ROV-mounted Full HD (1920 × 1080) resolution camera with a lighting system consisting of 16 bright LEDs in 4 × 4 stations. In total, coastal data consists of five 10 m transects: SM02-1, SM02-2, SM07-1, SM07-2 and SM08; (the latter being divided into two segments), while offshore data consists of two 30 s long video clips (DE01-1, DE01-2), which are accessible through the Mendeley cloud-based repository (Medelytė et al., 2022b).

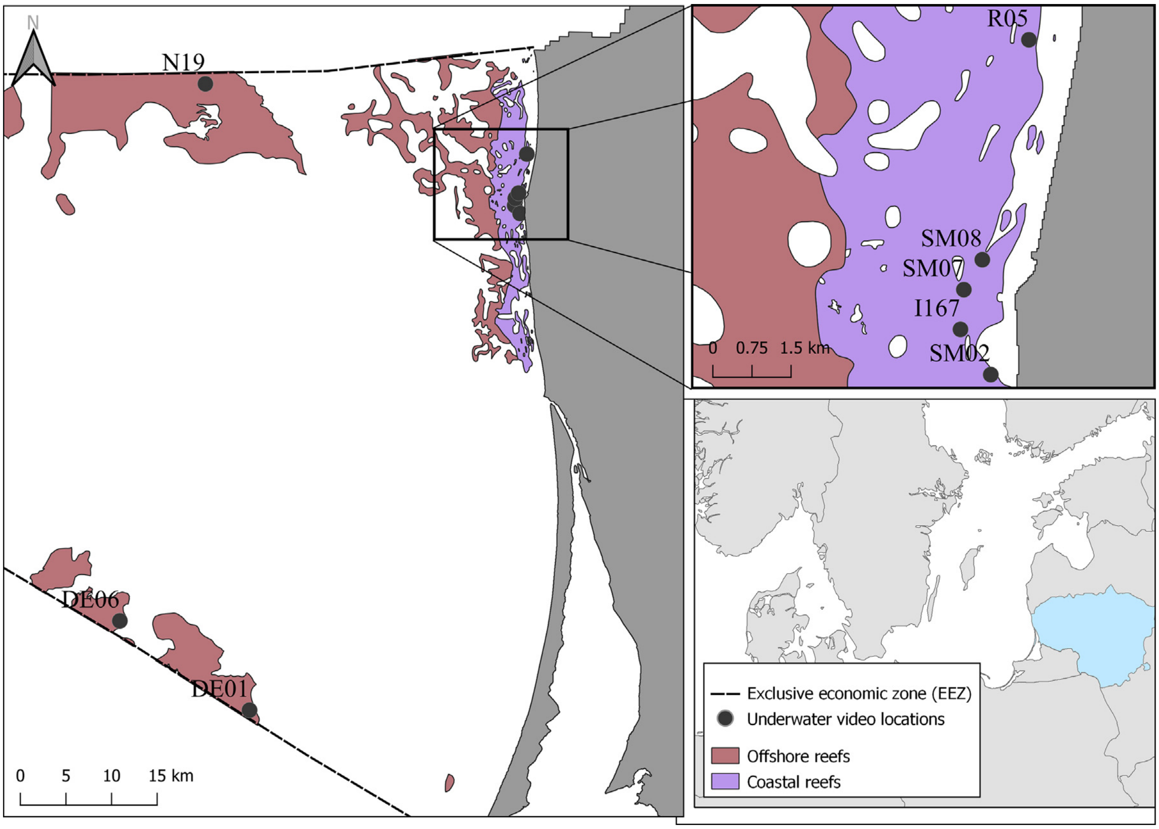

Figure 1 Underwater video sampling sites in South-Eastern Baltic Sea reefs.

For the validation of models, additional video data from four transects was used: R05, DE06, I167 and N19.

Ten video mosaics were created using a method developed by Rzhanov and Mayer (2004) while following steps outlined by Šaškov et al. (2015) and Šiaulys et al. (2021). The underwater mosaicking process is not always possible due to difficult weather conditions in open seas, since the "drop-down" camera is lifted by waves, resulting in an unstable camera distance from the bottom and thus complicating the frame-to-frame pairwise registration process needed for smooth mosaic construction. To address this issue, an experiment was carried out to test how the accuracy of biological and geological feature extraction changes when analysing mosaics and individual frames, i.e., whether it is possible to avoid the mosaicking step for accurate image analysis by analysing only frames. The frames were selected in such a way that adjacent frames did not overlap but were not too far apart (Figure 2), resulting in 148 extracted frames.

2.2. Extraction of frames

For the frame sampling approach, experts assigned a specific number of frames, typically 12—16, for each video transect, and equally spaced frames of 960 × 540 size were extracted using a command-line ffmpeg tool (Tomar, 2006). Additionally, seeking to obtain better quality representative frames, a complex sampling strategy was introduced: 1) Each video frame was converted to a high-dimensional feature vector of 1280 elements by using ImageNet pre-trained EfficientNet deep convolutional architecture (Tan and Le, 2019), the smallest and fastest EfficientNetB0 variant in the Python package image_embeddings. For example, DE01-2 video had 150 frames, and, after passing each frame through the model, we obtained a matrix of 150 × 1280 in size; 2) matrix obtained after converting frames to embeddings was further processed using sparse modelling for finding representative objects — Sparse Modelling Representative Selection (SMRS) algorithm (Elhamifar et al., 2012), using the authors' Python code. Parameters used: alpha=5, norm_type=2, thrS=0.99, thrP=0.98, max_iter=5000, step=100. This algorithm tries to find frames that are the most representative in a mathematical sense. 3) A representative frame that is closest to the frames selected by simple sampling is selected. Being close is defined here as smaller than 25% of the average distance between frames in simple sampling; 4) in a rare case, when no closest representative frames are detected, all frames around the corresponding frame (± average distance) from simple sampling are cut out into a smaller matrix, which is passed again to the SMRS algorithm, and steps 2—3 are repeated.

2.3. Manual annotation

All video mosaics and extracted frames of fixed size were annotated by 3 experts by drawing closed polygons (striving for pixel-level accuracy) using the Labelbox manual pixelwise segmentation tool (Labelbox). Dominant features of the SE Baltic coastal and offshore reefs were selected for

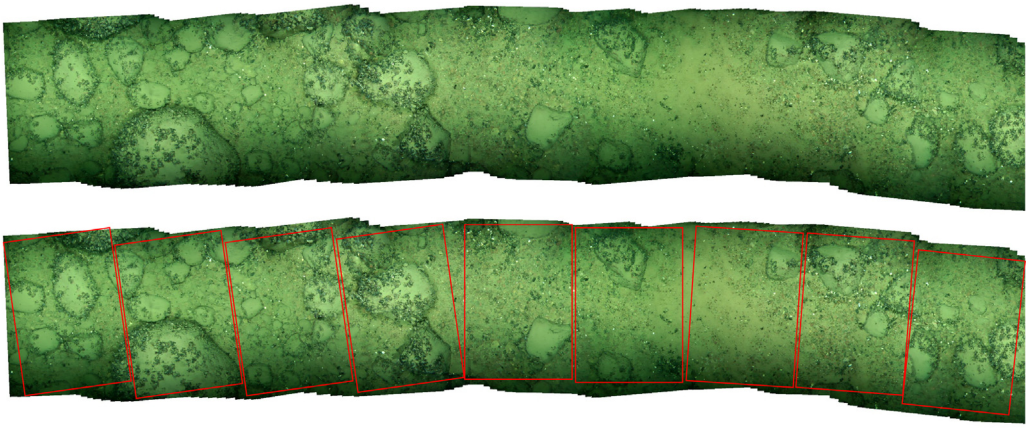

Figure 2 An example of a 2D mosaic and separate frames from the same video transect.

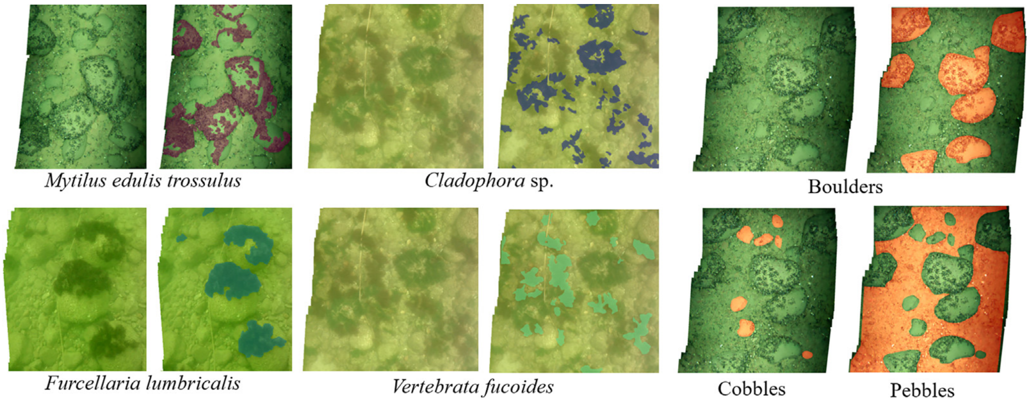

Figure 3 Annotated biological and geological features of SE Baltic coastal and offshore reefs.

annotation (see Figure 3). Selected biological features were red algae *Furcellaria lumbricalis* and *Vertebrata fucoides*, green algae *Cladophora* sp., blue mussel *Mytilus edulis trossulus*, geological features: boulders (>25 cm), cobbles (6—25 cm), pebbles (0.2—6 cm), and sand (<0.2 cm) according to the Wentworth scale (Wentworth, 1922). The summary of UI with mosaic sizes, number of frames, visual and modelled features is given in Table 1.

2.4. Pre-processing of underwater imagery

For the deep learning model with convolutional architecture, training and testing data were made by patching together underwater images that were either in the form of large mosaics or representative frames. This was done using the sliding window principle. Training patches were augmented to maximize the amount of information available and to provide a simple form of regularization. For evaluation, the transects were split in half to achieve 2-fold cross-validation.

Due to the limitations of the available computational resources, both mosaics and frames were sliced into overlapping 288 × 288-size patches. Overlap was the result of a sliding window or block processing idea with vertical and horizontal strides of 144 pixels. Due to the fact that mosaics contained a lot of white pixels, as a result of the mosaicking process, only patches with a minimum of 70% non-white pixels were considered as input images. Additionally, to increase the amount of training data, a few traditional augmentation techniques, such as vertical and horizontal flip, and one marine-specific technique, removal of water scattering (RoWS) (Chao and Wang, 2010), were used on the prepared input image patches.

2.5. Deep learning model for semantic segmentation

In the experiments, we used a deep convolutional neural network with pyramid spatial pooling architecture — the PSPNet model (Zhao et al., 2017) — with ImageNet pre-trained ResNet-34 (He et al., 2016) as the backbone. The PSPNet architecture takes its name from the so-called Pyramid Pooling Module, which helps the model capture the global context within the segmented image, leading to more successful pixel annotations using global information present in the image (Figure 4). In a nutshell, this module captures different resolutions of the feature map, trying to identify and preserve the most important features from the feature map (output from the backbone model), combining both the downsampled, convoluted, and upsampled features and the original feature map (obtained from the backbone model) itself. The model was implemented using the Keras framework (version 2.3.1), running on the Tensorflow backend (version 2.1.0), with the

Table 1 Summary of the underwater imagery. MP corresponds to the size of imagery in megapixels. The last 4 transects were only for a final hold-out validation. Modelled features were used for deep learning experiments.

Transect	Mosaic size	Mosaic MP	Frames	Frames MP	Modelled features	Additional features
SM02-1	3671 × 8285	8.20	16	8.29	Boulders	Cladophora, Vertebrata, Cobbles, Pebbles, Sand
SM02-2	4693 × 7307	8.36	14	7.26	Boulders	Cladophora, Vertebrata, Cobbles, Pebbles, Sand
SM07-1	2434 × 8774	9.04	16	8.29	Furcellaria, Boulder	Cobbles, Pebbles, Sand
SM07-2	5021 × 5107	7.06	12	6.22	Furcellaria, Boulder	Cobbles, Pebbles, Sand
SM08-1	4191 × 5379	6.64	12	6.22	Furcellaria, Boulder	Cladophora, Vertebrata, Cobbles, Pebbles, Sand
SM08-2	4745 × 5379	6.85	12	6.22	Furcellaria, Boulder	Cladophora, Vertebrata, Cobbles, Pebbles, Sand
DE01-1	1580 × 5480	5.17	11	5.70	Mytilus, Boulder	Cobbles, Pebbles, Sand
DE01-2	2434 × 8774	6.56	11	5.70	Mytilus, Boulder	Cobbles, Pebbles, Sand
DE06-1	1495 × 7087	6.87	10	5.18	Mytilus, Boulder	Cobbles, Pebbles, Sand
R05-1	1656 × 7113	7.11	9	4.67	Furcellaria, Boulder	Cobbles, Pebbles, Sand
I167	—	—	15	7.78	Furcellaria, Boulder	Cobbles, Pebbles, Sand
N19	—	—	10	5.18	Mytilus, Boulder	Cobbles, Pebbles, Sand

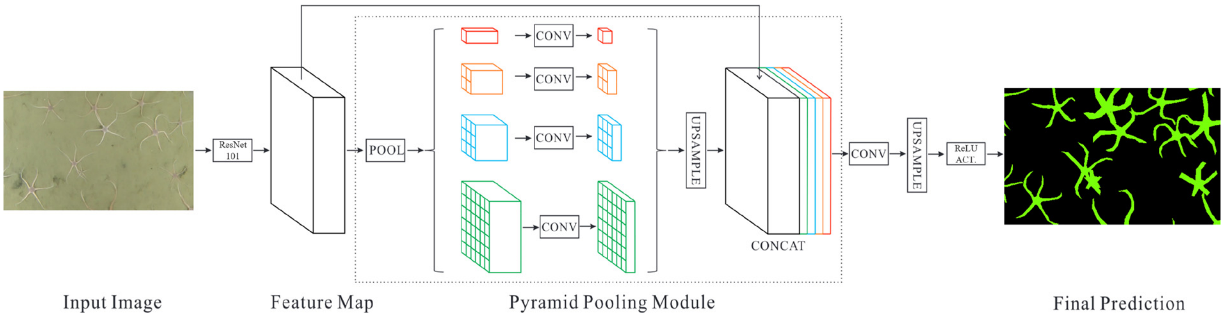

Figure 4 PSPNet model used for semantic segmentation. From Zhao et al. (2017) and Buškus et al. (2021).

help of the segmentation-models package (version 1.0.1) (Yakubovskiy, 2019). The models were trained for 50 epochs, with a batch size of 8 image patches.

2.6. Evaluation of model-based semantic segmentation

For training and testing the convolutional neural network model, we employed 2D mosaics and representative transect frames containing two biological and one geological feature (*F. lumbricalis*, *M. edulis trossulus*, and Boulders). The semantic segmentation task here was solved separately for each feature in a detection fashion. After a summary of manual annotation, the semantic segmentation task was evaluated by three types of validation schemes: 1) 2-fold transect-stratified cross-validation where each transect was split in half and either all bottom parts or all top parts of transects were used for training; 2) leave one transect out validation where a single transect is used for testing while training on all the remaining transects; 3) hold-out validation had additional unseen imagery with features of interest collected and annotated as a way to stress-test the semantic segmentation task.

Stratification by transect in a 2-fold CV means that each transect is split in half -top and bottom parts — and training is performed on one part while testing on the other part. For example, after training on all the bottom parts of mosaics (or the first half of the corresponding frame set), testing is performed on all the top parts, and vice versa. Such a strategy guarantees that testing is performed on somewhat similar imagery to the one the model was trained on. However, the drawback is that smaller amounts of training data

(50/50% split instead of a more common 80/20%) and the count of feature instances (objects of interest) can differ to a large extent between the top and bottom parts of the transect.

The main benefit of the leave-one-out validation (LOO) strategy is the use of all available training data, but the drawback is the lack of stratification by transect, where in this strategy it is designed fully as the testing data. In this kind of validation model, usefulness can be fully investigated, but the testing data can differ from training due to visual differences between transects.

Due to the selected transects for the testing split, the hold-out validation strategy was the most challenging validation. The transects were recorded at different times and using different video recording equipment. Moreover, two of the four transects could not be stitched into mosaics because of poor image quality resulting from strong waves at the recording time. For those two challenging transects (I167 and N19), only the selection of representative video frames was possible.

The success of segmentation was determined by the intersection over union (IOU) metric, and estimates of visual coverage were calculated. This is a common metric in semantic image segmentation (Elbode et al., 2020), measuring segmentation success by comparing the ground truth with the prediction mask (that is, annotated and predicted image pixels), also known as the Jaccard index. The metric is defined as:

$$IOU = \frac{true\ positive}{true\ positive + false\ negative + false\ positive}$$

In addition, final prediction masks were used to estimate the visual coverage of the feature in question. The coverage itself was interpreted as a ratio between predicted or ground truth masks (that is, 'active' pixels) with only relevant pixels (excluding white pixels), in the mosaic setting and all pixels in the frame setting.

3. Results

Distinct benthic communities represented the sites chosen for this study. Coastal sites (SM) were dominated by macroalgae, while offshore sites (DE) were dominated by mussels (Figure 5). Three coastal sites were also different: The shallowest (4 m) SM02 site was dominated by the green algae *Cladophora* sp. (23.1±0.5%) and red algae *V. fucoides* (14.5±1.3%), with only a few thalli of *F. lumbricalis* (0.1%). On the contrary, the SM08 site was dominated by *F. lumbricalis* (49.4±10.1%) with only a few *Cladophora* sp. (3.1±3.2%) and *V. fucoides* (8.1±7.7%). At the SM07 site, only scarce patches of *F. lumbricalis* were present (10.8±1.5%). The substrate in all sites was dominated by coarse sediments: boulders (32.4—75.1%), cobble (2.8—16.8%), pebble (10.7—54.4%), while the sand fraction had the lowest share (7.8±6.2%).

3.1. Comparison of manual expert annotations

A comparison of 2D mosaic versus sampled frames with respect to expert-based manual annotations was done first. The possibility of sparsifying selected frames and using fewer images is evaluated with respect to coverage estimates. The accuracy was measured as the absolute difference from the mosaic coverage estimates (baseline) according to the following heuristic scale: 0—5% excellent, 5—10% good, 10—20% moderate, >20% bad.

Table 2 Average differences (± standard deviation) between expert coverage estimations of 2D mosaics (baseline) and frames for biological and geological visual features from all samples.

Feature	All frames	1/2 frames	1/3 frames
F. lumbricalis	4±2.8	3.9±2.2	5.5±3
M. edulis trossulus	1.6±1.2	3.3±2.1	1.7±2.2
Cladophora sp.	1.3±1	1.2±0.5	1.4±0.7
V. fucoides	2.7±3.6	2.7±3.5	4.4±5.6
Boulders	4.9±7.3	5.6±6.8	5.4±4.1
Cobble	2.1±1.9	2.2±1.8	2.8±1.5
Pebble	5.8±4.8	6±4.7	6.1±3.5
Sand	2.8±3.7	2.8±2.5	3.7±6.1

As explained in Section 2.1, two methods for frame extraction from videos were used: simple and complex. For each video transect, we estimated the differences in coverage estimations (both for biological and geological visual features) from the baseline (mosaics) for both methods. The pairwise Wilcoxon signed-rank test showed that the simple method gives significantly lower baseline differences than the complex method (test statistic = 2.31 with continuity correction applied, p-value = 0.01). Thus, to make further analysis less complicated, only estimations of simple frame selection are provided further in this subsection.

The results have shown that, in general, the quality of visual evaluation does not suffer when analysing individual frames when comparing the differences between mosaics and frames (Table 2). Analysis of all frames had excellent accuracy and differed less than 5% from the baseline, except pebble (5.8%, good accuracy); excellent accuracy was also achieved from the analysis with a reduced number of frames, with the exception of boulders, which differed 5.4—5.6% from the baseline (good accuracy).

The differences from the baseline *F. lumbricalis* coverage estimates from all frames ranged from 1.3 to 7.0% in individual mosaics, indicating excellent to good accuracy. The reduction of frames provided similar results (Table 3). It is noticeable that in transects with lower coverage of *F. lumbricalis*, a higher accuracy was reached: transects with <20% coverage maintained excellent accuracy even with frame reduction, while transects with >40% coverage showed lower but still good accuracy.

The accuracy of boulder estimation from all frames was excellent in 6 out of 8 transects; in the other two transects from SM02 site, experts significantly overestimated the boulder class, resulting in moderate and bad accuracies (Table 4). The reduction of analysed frames resulted in similar accuracy in all transects except SM08-1, where the analysis of 1/3 frames resulted in increased differences from the baseline from 1.9 to 9.2%.

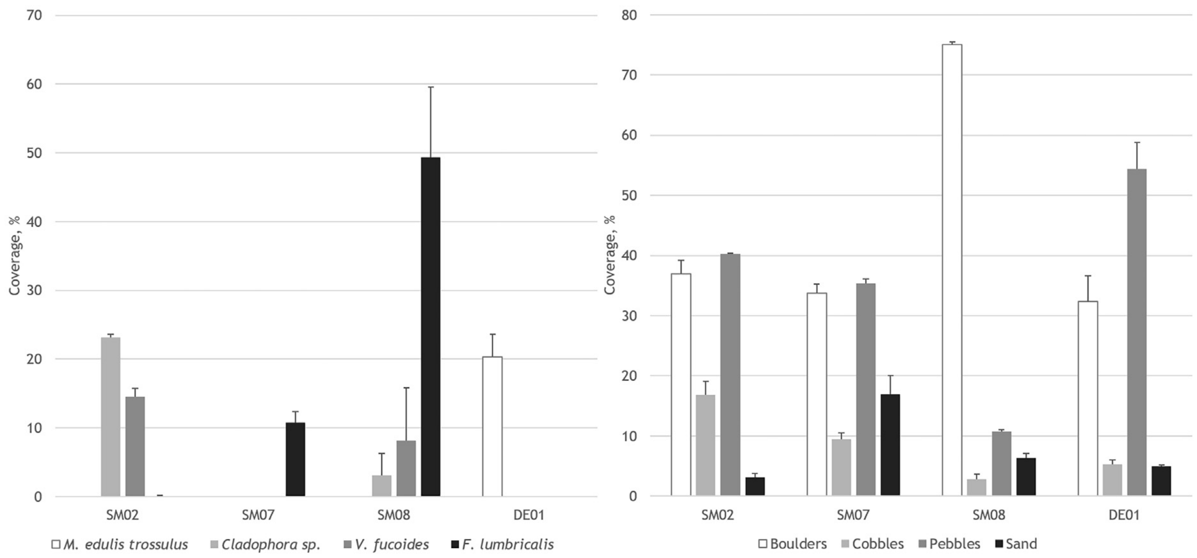

Figure 5 The percentage coverage of biological (*Furcellaria lumbricalis*, *Vertebrata fucoides*, *Cladophora* sp., *Mytilus edulis trossulus*) and geological (boulders, cobbles, pebbles and sand) visual features at research sites.

Table 3 Seabed visual coverage estimates (in %) of *F. lumbricalis* from the analysis of mosaics, all frames, half of frames and one third of frames. Note: Δ — differences from the mosaics (baseline).

Furcellaria transects	Mosaic	All frames	Δ	1/2 frames	Δ	1/3 frames	Δ
				Coverage, %			
SM07-1	9.7	8.4	1.3	8.9	0.8	7.3	2.4
SM07-2	11.8	10.0	1.9	7.9	3.9	8.4	3.5
SM08-1	42.2	47.8	−5.6	47.6	−5.4	50.1	−7.9
SM08-2	56.6	49.5	7.0	51.1	5.4	48.4	8.2

Table 4 The coverage (%) of boulders from the analysis of mosaics, all frames, half of frames and one-third of frames. Note: Δ — differences from the mosaics (baseline).

Boulder transects	Mosaic	All frames	Δ	1/2 frames	Δ	1/3 frames	Δ
				Coverage, %			
SM02-1	38.6	50.5	−11.9	50.6	−12.1	44.5	−5.9
SM02-2	35.5	55.8	−20.3	55.3	−19.8	48.8	−13.3
SM07-1	34.7	35.0	−0.2	32.4	2.4	29.7	5.1
SM07-2	32.6	32.4	0.2	31.8	0.8	34.7	−2.1
SM08-1	74.8	77.4	−2.7	80.0	−5.3	70.4	4.3
SM08-2	75.3	77.2	−1.9	74.4	0.9	84.5	−9.2
DE01-1	35.4	37.4	−2.0	37.5	−2.1	36.8	−1.4
DE01-2	29.4	29.8	−0.4	27.8	1.6	31.4	−2.0

3.2. Two-fold transect-stratified cross-validation

The segmentation success for the *F. lumbricalis* feature was very good, with the resulting IOU score in the 0.611—0.839 range (Table 5). Interestingly, the IOU score using frames was often higher than using mosaic, by 0.035 overall. Successful segmentation also resulted in accurate seabed visual coverage estimates (Table 6), where the estimate from expert annotations was 28.81%, the prediction of the model trained in mosaics was 28.57%, and the predictions from models trained on frames were 27.55% and 27.92%. Individually, for separate transects, the difference between expert estimates on mosaic and models trained on frames did not exceed 8.5 percentage points.

The segmentation success for the *M. edulis trossulus* feature was moderate, with the resulting IOU score in the 0.560—0.699 range (Table 5). Frames performed similarly to mosaic with overall differences of —0.051 and 0.005 in the IOU score, but a simple sampling of frames provided more accurate coverage estimates. Poorer segmentation

Table 5 Segmentation performance, as measured by IOU score, using 2-fold and LOO validation for *F. lumbricalis, M. edulis trossulus* and boulders features in mosaic imagery or selected representative video frames (by simple or complex sampling).

		Mosaic		Simple		Complex	
		IOU		△ IOU		△ IOU	
Feature	Transect	2-fold	LOO	2-fold	LOO	2-fold	LOO
Furcellaria lumbricalis	SM07-1	0.703	0.694	−0.015	−0.021	0.016	−0.043
	SM07-2	0.839	0.799	0.046	0.035	0.030	0.070
	SM08-1	0.661	0.628	−0.104	−0.127	−0.050	−0.099
	SM08-2	0.726	0.665	0.000	−0.047	0.028	−0.039
	Totals:	**0.711**	**0.664**	**−0.035**	**−0.077**	**−0.035**	**−0.054**
Mytilus edulis trossulus	DE01-1	0.671	0.486	−0.028	−0.113	0.064	−0.006
	DE01-2	0.560	0.600	−0.061	0.003	−0.050	0.005
	Totals:	**0.613**	**0.549**	**−0.051**	**0.005**	**0.005**	**0.010**
Boulders	DE01-1	0.670	0.551	−0.063	−0.112	0.012	−0.082
	DE01-2	0.649	0.601	−0.085	−0.074	−0.024	0.028
	SM07-1	0.582	0.566	−0.064	−0.024	0.060	0.022
	SM07-2	0.517	0.434	−0.111	−0.178	0.039	−0.015
	SM02-1	0.344	0.430	−0.256	−0.183	−0.091	−0.052
	SM02-2	0.297	0.279	−0.325	−0.264	−0.057	−0.117
	SM08-1	0.806	0.785	−0.016	−0.042	0.041	0.018
	SM08-2	0.790	0.776	−0.046	−0.057	0.014	−0.006
	Totals:	**0.598**	**0.578**	**−0.108**	**−0.102**	**−0.003**	**−0.025**

Table 6 Seabed visual coverage estimates, as measured in percentages, using 2-fold and LOO validation for *F. lumbricalis, M. edulis trossulus* and boulders features in mosaic imagery or selected representative video frames (by simple or complex sampling). Abbreviations: GT (ground-truth) — results of expert annotations; △ DL — difference of model-based predictions (DL) from mosaic-wise ground-truth annotations (GT—DL).

		Mosaic			Simple		Complex	
		GT	△ DL		△ DL		△ DL	
Feature	Transect		2-fold	LOO	2-fold	LOO	2-fold	LOO
Furcellaria lumbricalis	SM07-1	9.75	0.31	−0.23	0.74	0.48	1.10	2.09
	SM07-2	11.91	0.86	1.26	2.66	3.11	1.45	2.89
	SM08-1	43.15	−0.68	−17.88	−5.89	−9.17	−5.19	−8.46
	SM08-2	57.48	0.41	13.07	8.41	6.82	6.82	4.42
	Totals:	**28.81**	**0.24**	**−0.75**	**1.26**	**0.16**	**0.89**	**0.22**
Mytilus edulis trossulus	DE01-1	22.83	0.76	9.92	3.62	7.59	7.07	11.18
	DE01-2	18.11	0.20	−0.82	1.86	−1.14	4.78	1.73
	Totals:	**20.19**	**0.44**	**3.91**	**2.46**	**2.94**	**5.64**	**6.17**
Boulders	DE01-1	29.65	4.09	7.87	−0.28	2.60	3.91	1.76
	DE01-2	35.60	2.62	5.73	1.49	2.14	4.44	0.18
	SM07-1	34.95	0.55	4.47	1.99	1.05	10.72	9.88
	SM07-2	32.82	9.36	13.74	3.28	5.76	15.30	16.00
	SM02-1	38.80	10.42	−4.85	−24.13	−24.14	3.76	−4.53
	SM02-2	35.40	8.57	13.56	−22.95	−11.66	10.78	6.15
	SM08-1	76.41	−3.59	−0.86	−6.59	−3.30	−0.53	−1.30
	SM08-2	76.55	−2.70	3.86	−7.70	−5.17	2.01	2.57
	Totals:	**44.37**	**3.99**	**5.23**	**−7.72**	**−5.02**	**6.42**	**3.77**

accuracy did not noticeably affect seabed visual coverage estimates (Table 6), where the estimate of expert annotations was 20.19%, prediction from model trained on mosaics was 19.75% and predictions from models trained on frames were 17.73% and 14.55%. Individually, for separate transects, the difference between the expert estimate on the mosaic and the model trained on frames (from simple sampling) did not exceed 4 percentage points.

The segmentation success of the boulder feature was very varied, with much better results for simple frame sampling than using mosaics, with the resulting IOU score in the range 0.297—0.837 range (Table 5). Simple sampling

Table 7 Segmentation performance, as measured by IOU score, using hold-out validation for *F. lumbricalis*, *M. edulis trossulus* and boulders features in mosaic imagery or selected representative video frames by simple sampling.

		Mosaic	Frames	
Feature	*Transect*	IOU	IOU	△ IOU
Furcellaria lumbricalis	R05-1	0.259	0.258	0.001
	I167	—	0.824	—
Mytilus edulis trossulus	DE06-1	0.061	0.075	−0.014
	N19	—	0.441	—
Boulders	R05-1	0.161	0.349	−0.188
	DE06-1	0.148	0.345	−0.197
	I167	—	0.453	—
	N19	—	0.143	—

provided the best segmentation accuracy overall, with an IOU score of 0.706. We suspect that such differences in IOU could be due to many objects in boulder class having poor visibility in the SM02_1 and SM02_2 mosaics, which also resulted in significant differences from expert annotations when using deep learning model predictions. The visual coverage estimates were markedly affected (Table 6), especially for SM02 and SM07 transects, where the overall estimate of the expert annotations was 44.37%, the prediction of the model trained on mosaics was 40.38% (underestimate of ~4 percentage points), and the predictions from models trained on frames were 52.09% (overestimate of ~7.7 percentage points) using simple and 37.95% (underestimate of ~6.4 percentage points) using complex frame sampling. Excluding too large overestimates of visual coverage using simple sampling of SM02 frames and underestimates using complex sampling of SM07-2 frames individually for separate transects, the difference between the expert estimate on the mosaic and model predictions did not exceed 11 percentage points.

3.3. Leave one transect out validation

The segmentation success for the *F. lumbricalis* feature was very good, with the resulting IOU score in the 0.628—0.799 range (Table 5). The IOU score was again higher for the *F. lumbricalis* feature using simple and complex frames sampling than using mosaic (overall by 0.077 and 0.054 respectively). Successful segmentation also resulted in accurate seabed visual coverage estimates (Table 6), where the estimate from expert annotations was 28.81%, the prediction of the model trained in mosaics was 29.56%, and predictions from models trained on frames were 28.65% and 28.59%. Individually, for separate transects, the difference between the expert estimate on the mosaic and the model trained on frames did not exceed 9.2 percentage points.

The segmentation success for the *M. edulis trossulus* feature was moderate, with the resulting IOU score in the 0.486—0.600 range (Table 5). Frames performed similarly to mosaic, with overall differences of 0.005 and 0.01 in the IOU score, but mosaic provided slightly more accurate coverage estimates. Poorer segmentation accuracy did not noticeably affect seabed visual coverage estimates (Table 6), where the estimate of expert annotations was 20.19%, prediction from model trained on mosaics was 16.28% and predictions from models trained on frames were 17.25% and 14.02%. Individually, for separate transects, the difference between expert estimates on mosaic and models trained on frames did not exceed 11.2 percentage points.

The segmentation success of the boulder feature was very varied, with much better results for simple frame sampling than using mosaics, with the resulting IOU score in the 0.279—0.833 range (Table 5). Simple sampling provided the best segmentation accuracy overall, with an IOU score of 0.680. Similarly, for 2-fold CV results, we hypothesize that significant discrepancies in IOU may be related to the low visibility of several boulder-class objects in SM02-1 and SM02-2 mosaics, which also led to large differences between expert annotations and deep learning model predictions. The visual coverage estimates were markedly affected (Table 6), especially for SM02 and SM07 transects, where the overall estimate from expert annotations was 44.37%, the prediction from model trained in mosaics was 39.13% (underestimate of ~5.2 percentage points), and the predictions from models trained on frames were 49.38% (overestimate of ~5 percentage points) using simple and 40.60% (underestimate of ~3.8 percentage points) using complex frame sampling. Excluding too large overestimates of visual coverage using simple sampling of SM02 frames and underestimates using complex sampling of SM07-2 frames individually for separate transects, the difference between the expert estimate on the mosaic and model predictions did not exceed 16 percentage points.

3.4. Stress testing with hold-out validation

The segmentation success for the *F. lumbricalis* feature was poor for the R05-1 and excellent for the I167 transect (Table 7). Despite such different results, seabed visual coverage estimates were of acceptable accuracy (Table 8), deviating from ground-truth expert annotations by 6.7 percentage points for mosaic and just 3.4 or 0.93 percentage points for frames. Surprisingly, the frames outperformed the mosaic for the R05-1 transect. For comparison, the difference between expert annotations of mosaic and frames was 1.7 percentage points.

The segmentation success for the *M. edulis trossulus* feature was unacceptable for the DE06-1 and mediocre for the N19 transect (Table 7). Due to the low visual coverage in the DE06-1 transect (Table 8), where experts estimated

Table 8 Seabed visual coverage estimates, as measured in percentages, using hold-out validation for *F. lumbricalis, M. edulis trossulus* and boulders features in mosaic imagery or selected representative video frames by simple sampling. Abbreviations: GT (ground-truth) — results of expert annotations; DL (deep learning) — results of model-based predictions; Δ — difference from mosaic-wise ground-truth annotations (GT—DL). In case a mosaic was not available frame-based GT annotations were used.

		Mosaic		Frames			
Feature	Transect	GT	Δ DL	GT	Δ GT	DL	Δ DL
Furcellaria	R05-1	10.18	6.70	8.48	1.70	6.78	3.40
lumbricalis	I167	—	—	40.46	—	39.53	0.93
Mytilus edulis	DE06-1	7.76	7.24	8.40	−0.64	0.66	7.11
trossulus	N19	—	—	33.82	—	25.59	8.23
Boulders	R05-1	16.19	−51.23	32.13	−15.94	83.68	−67.49
	DE06-1	24.84	17.59	24.62	0.22	15.40	9.44
	I167	—	—	44.59	—	50.30	−5.71
	N19	—	—	51.09	—	32.13	18.96

7.76% in mosaics and 8.4% in frames (with a small overestimate of 0.64 percentage points), differences of ~7 percentage points are too large in this case. Basically, this means that the model could not predict lower amounts of *M. edulis trossulus* feature objects in the DE06-1 transect when trained on DE01-2 and DE01-1 transects, which had higher amounts. Meanwhile, the more successfully segmented N19 transect frames with higher amounts of *Mytilus* features had a seabed visual coverage estimate of 33.82% by experts and a 25.59% underestimate by the model (Table 8), where the difference of ~8 percentage points can be seen as a good result.

The segmentation success of the boulder feature was poor regardless of the transect but, interestingly, much better for frames than for mosaics (Table 7). However, the visual coverage for the R05-1 transect was unacceptably overestimated (Table 8), with the differences between the expert and the model being too large. Meanwhile, other transects showed better results, with the largest difference being ~19 percentage points.

5. Discussion

Our results have shown that in general, coverage estimations from mosaics and frames were very similar for all eight features, thus providing a few opportunities for more effective UI analysis. Using a set of representative frames from an underwater video may considerably reduce the time required for preprocessing raw data since mosaicking of seabed imagery can be labour-intensive and requires specific software or algorithms and professional knowledge, despite existing tools such as AutoStitch, APAP, and SPHP (Li et al., 2019). The most time-consuming step of mosaicking is manual registration of consecutive frames if automatic pair-wise registration of these frames is unsuccessful. This is often the case for Baltic Sea UI which is usually of limited quality (high turbidity, camera motion due to waves, motion of features, changing lighting, etc.). Also, the requirement of irregular manual intervention to mosaicking process makes a fully automated video analysis very complicated. Frames-based approach with a decreased number of analysed frames also provides a reasonable option to reduce the efforts needed for UI annotation, considering that a single 100 m video transect, depending on the drift, can result in 100—150 individual frames. However, this approach has some implicit limitations. While the frame-based analysis may well substitute mosaicking for features that require coverage estimation (such as underwater vegetation, colonial fauna, and substrate types), this approach is less successful in estimating the number of individual organisms, especially if they are rare and scarce or moving during the video. We noticed that some individuals can be counted twice if they are partially annotated in two adjacent frames, or cannot be counted at all if they are located between adjacent frames (Figure 6). Furthermore, reducing the number of frames analysed in this case may lead to significant overestimations or underestimations, depending on whether or not a rare feature occurs in the analysed set of frames.

Surprisingly, complex sampling did not provide a clear advantage over simple sampling, with semantic segmentation performance always inferior (resulting in lower IOU values) and coverage estimations depending on the class analyzed (only with marginally better results for *F. lumbricalis* and boulder classes). Due to poorer segmentation results and negligible differences in coverage estimates, the introduced complex sampling cannot be recommended since it requires a large computational overhead while not providing better results. One of the reasons why complex sampling failed could be the high dimensionality of EfficientNet embeddings applied to short video segments, creating a curse of dimensionality challenge (Moghaddam et al., 2020) for the underlying SMRS algorithm.

Semantic segmentation gave moderate accuracy, as measured by the IOU score, but seabed coverage estimates obtained from predicted segmentations were overall quite accurate. Based on leave-one-out and two-fold transect stratified validations, the total absolute differences between expert annotations and model results were less than 6% and 8%, respectively, which is quite impressive considering the often higher variability of intra- and interobserver classification (Beijbom et al., 2015; Reeves et al., 2007). However, some individual transects resulted in high biases of up to 24% in absolute difference.

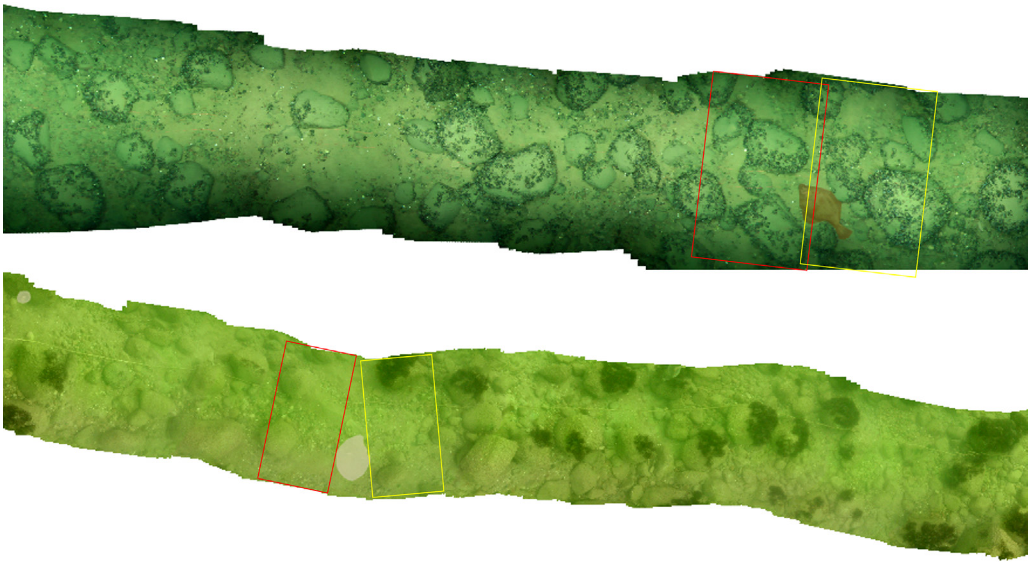

Figure 6 Double annotation of European flounder *Platichthys flesus* in two adjacent frames (top picture, DE01-1 transect) and misannotation of moon jelly *Aurelia aurita* between adjacent frames (bottom picture, SE07-1 transect).

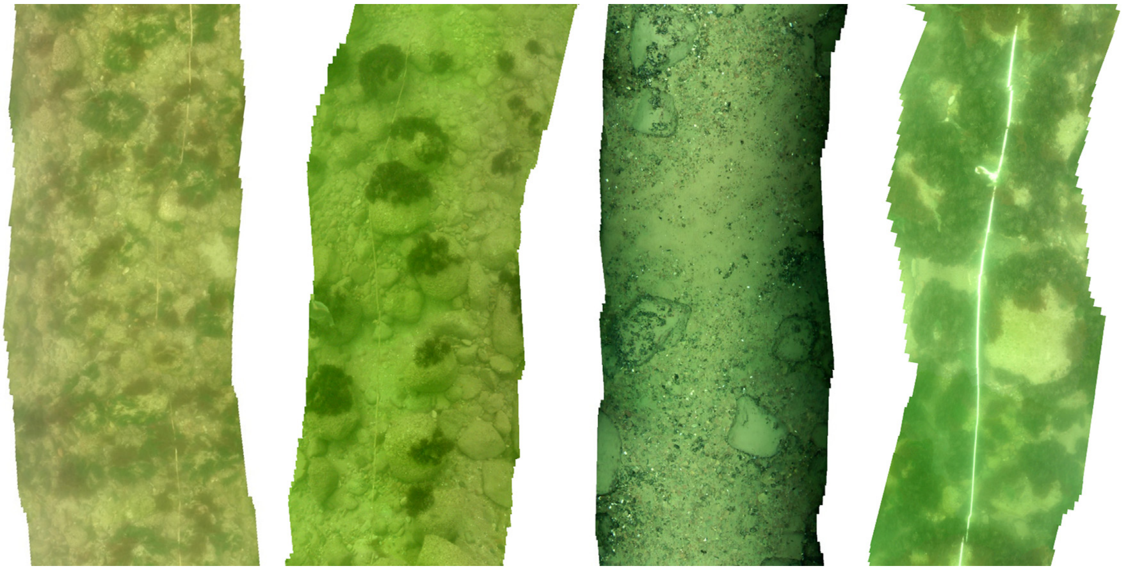

Figure 7 The performance of deep learning models for the boulder class based on IOU scores from worst (left) to best (right) in SM02-1, SM07-1, DE01-2 and SM08-2 transects respectively.

The validation of the boulder class gave somewhat unexpected results. The best IOU scores were for SM08 transects where boulders were mostly overgrown by macroalgae *F. lumbricalis* and with hardly visible outlines, while transects with relatively easily outlined boulders gave lower IOU scores (Figure 7). This could be explained by the substrate preferences of different macroalgae species. For example, as stated by Bučas et al. (2007), in Lithuanian coastal reefs, perennial red algae *F. lumbricalis* (the dominant feature of SM08 transects) prefer the most stable substrate — boulders, while green algae *Cladophora* sp. (the dominant feature of SM02 transects) can overgrow both boulders and cobbles. This could suggest that during substrate classification, the model considers epibenthos and tends to assign overgrown substrate to the boulder class rather than to cobble, whereas, in transects with scarce vegetation (SM07) or vegetation both on boulders and cobbles (SM02), the classification is less accurate. On the other hand, model results from the frame analysis were more accurate than from mosaics, showing that the framing approach is not only more effective for the annotation of UI but also more suitable for deep learning models.

Stress tests with hold-out validation, which were based on additionally annotated challenging test data, resulted in even worse model performance, with differences between experts and the model exceeding 50% for some transects. This could be explained by intentionally selecting videos with different image quality for the test dataset. For example, I167 and N19 sites were filmed with a "drop-down" video system with noticeable wave action, while the light environment and image sharpness in R05-1, sediment composition, and colour palette in DE06-1 were also different. This emphasizes the importance of having a training dataset with a variety of filming equipment and environmental conditions. This seems especially important for the very dynamic environment of Lithuanian coastal reefs, which are under the influence of plume from the Curonian Lagoon (Vaičiūtė et al., 2012), regular upwelling events (Dabuleviciene et al., 2018), waves and currents, not to mention different cloudiness, all of which can determine different lighting, water colour, transparency/turbidity, camera motion, and other parameters that can influence the results of visual analysis.

6. Conclusions

Our study has shown that seabed coverage estimations from video mosaics and individual frames provided similar results, suggesting that the mosaicking step, often used for UI analysis, could be skipped if an approximate estimate of biological and geological features is sufficient. Moreover, results indicated that even a two- or three-fold decrease in the frames analysed still resulted in relatively accurate coverage estimates for most of the features. In general, coverage estimates from automatic segmentation with deep learning models gave very promising seabed coverage estimation results for all visual classes, despite moderate IOU scores. Frame-based results were often slightly worse than mosaic-based results, but these differences seem to be negligible. When comparing seabed visual coverage estimates from expert annotated mosaics with the estimates from model-based segmentation predictions, absolute differences did not exceed 11% in 2-fold transect-stratified cross-validation, 16% in the leave-one-transect-out validation scheme, and 19% in challenging hold-out validation overall. Interestingly, the largest differences were consistently obtained for the boulder feature, which had large percentages of objects, resulting in large visual coverage. Judging from observed biases differing with respect to the validation scheme, we could advise having more varied imagery, both in recording equipment and environmental conditions. Therefore, contrary to coverage estimates from expert annotation of frames recommendations, we do not recommend reducing the number of selected frames if the goal is to prepare underwater imagery for deep learning model training. Finally, this study has laid a solid stepping stone towards automatic recognition and estimation of SE Baltic hard bottom features from UI, which in the future could considerably facilitate reef monitoring and environmental status assessment.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by the project DEMERSAL "A deep learning-based automated system for seabed imagery recognition" funded by the Research Council of Lithuania under the agreement No. P-MIP-19-492.

References

Alonso, I., Yuval, M., Eyal, G., Treibitz, T., Murillo, A.C., 2019. CoralSeg: Learning coral segmentation from sparse annotations. J. Field Robot. 36 (8), 1456—1477. https://doi.org/10.1002/rob.21915

Beijbom, O., Edmunds, P.J., Roelfsema, C., Smith, J., Kline, D.I., Neal, B.P., Dunlap, M.J., Moriarty, V., Fan, T.Y., Tan, C.J., Chan, S., 2015. Towards automated annotation of benthic survey images: Variability of human experts and operational modes of automation. PLOS One 10 (7), e0130312. https://doi.org/10.1371/journal.pone.0130312

Bučas, M., Daunys, D., Olenin, S., 2007. Overgrowth patterns of the red algae *Furcellaria lumbricalis* at an exposed Baltic Sea coast: the results of a remote underwater video data analysis. Estuar. Coast. Shelf S. 75 (3), 308—316. https://doi.org/10.1016/j.ecss.2007.04.038

Buhl-Mortensen, L., Buhl-Mortensen, P., Dolan, M.F., Holte, B., 2015. The MAREANO programme—A full coverage mapping of the Norwegian off-shore benthic environment and fauna. Mar. Biol. Res. 11 (1), 4—17. https://doi.org/10.1080/17451000.2014.952312

Buškus, K., Vaičiukynas, E., Verikas, A., Medelytė, S., Šiaulys, A., Šaškov, A., 2021. Automated quantification of brittle stars in seabed imagery using computer vision techniques. Sensors 21 (22), 7598. https://doi.org/10.3390/s21227598

Casoli, E., Ventura, D., Mancini, G., Pace, D.S., Belluscio, A., Ardizzone, G., 2021. High spatial resolution photo mosaicking for the monitoring of coralligenous reefs. Coral Reefs 40 (4), 1267—1280. https://doi.org/10.1007/s00338-021-02136-4

Chao, L., Wang, M., 2010. Removal of water scattering. In: International conference on computer engineering and technology IEEE, 2, 35. https://doi.org/10.1109/ICCET.2010.5485339

Dabuleviciene, T., Kozlov, I.E., Vaiciute, D., Dailidiene, I., 2018. Remote sensing of coastal upwelling in the south-eastern Baltic Sea: Statistical properties and implications for the coastal environment. Remote Sens. 10 (11), 1752. https://doi.org/10.3390/rs10111752

Eelbode, T., Bertels, J., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., Blaschko, M.B., 2020. Optimization for medical image segmentation: theory and practice when evaluating with dice score or Jaccard index. IEEE T. Med. Imaging 39 (11), 3679—3690. https://doi.org/10.1109/TMI.2020.3002417

Elhamifar, E., Sapiro, G., Vidal, R., 2012. See all by looking at a few: Sparse modeling for finding representative objects. In: Proc. CVPR IEEE, 1600—1607. https://doi.org/10.1109/CVPR.2012.6247852

Ferreira, T., Rasband, W., 2012. ImageJ user guide — IJ 1.46. URL: https://imagej.net/docs/guide/ (accessed on 2 August 2023).

Gracias, N., Garcia, R., Campos, R., Hurtos, N., Prados, R., Shihavuddin, A.S.M., Nicosevici, T., Elibol, A., Neumann, L., Escartin, J., 2017. Application challenges of underwater vision. In: López, A.M., Imiya, A., Pajdla, T., Álvarez, J.M. (Eds.), Computer Vision in Vehicle Technology: Land, Sea & Air. Wiley, Chichester, 133—160. https://doi.org/10.1002/9781118868065.ch7

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. Proc. CVPR IEEE, 770—778. https://doi.org/10.1109/CVPR.2016.90

Islam, M.J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., Enan, S.S., Sattar, J., 2020. Semantic segmentation of underwater imagery: Dataset and benchmark. IEEE Int. C. Int. Robot. 1769—1776. https://doi.org/10.1109/iros45743.2020.9340821

Kohler, K.E., Gill, S.M., 2006. Coral Point Count with Excel extensions (CPCe): A Visual Basic program for the determination of coral and substrate coverage using random point count methodology. Comput. Geosci. 32 (9), 1259—1269. https://doi.org/10.1016/j.cageo.2005.11.009

Labelbox. URL: https://labelbox.com (accessed 2 August 2023).

Langenkämper, D., Zurowietz, M., Schoening, T., Nattkemper, T.W., 2017. Biigle 2.0-browsing and annotating large marine image collections. Front. Mar. Sci. 4, 83. https://doi.org/10.3389/fmars.2017.00083

Li, Y., Randall, C.J., van Woesik, R., Ribeiro, E., 2019. Underwater video mosaicing using topology and superpixel-based pairwise stitching. Expert Syst. Appl. 119, 171—183. https://doi.org/10.1016/j.eswa.2018.10.041

Liu, F., Fang, M., 2020. Semantic segmentation of underwater images based on improved Deeplab. J. Mar. Sci. Eng. 8 (3), 188. https://doi.org/10.3390/jmse8030188

Martin-Abadal, M., Guerrero-Font, E., Bonin-Font, F., Gonzalez-Cid, Y., 2018. Deep semantic segmentation in an AUV for online *Posidonia oceanica* meadows identification. IEEE Access 6, 60956—60967. https://doi.org/10.1109/ACCESS.2018.2875412

Medelytė, S., Šiaulys, A., Daunys, D., Włodarska-Kowalczuk, M., Węsławski, J.M., Olenin, S., 2022a. Application of underwater imagery for the description of upper sublittoral benthic communities in glaciated and ice-free Arctic fjords. Polar Biol. 45 (12), 1655—1671. https://doi.org/10.1007/s00300-022-03096-3

Medelytė, S., Šiaulys, A., Vaiciukynas, E., Buškus, K., Šaškov, A., Olenin, S., 2022b. A fully-annotated imagery dataset of sublittoral benthic species in South Eastern Baltic Sea reefs. Mendeley Data, V1. https://doi.org/10.17632/wsd42v8mk5.1

Moghaddam, S.H.A., Mokhtarzade, M., Beirami, B.A., 2020. A feature extraction method based on spectral segmentation and integration of hyperspectral images. Int. J. Appl. Earth Obs. 89, 102097. https://doi.org/10.1016/j.jag.2020.102097

Piechaud, N., Hunt, C., Culverhouse, P.F., Foster, N.L., Howell, K.L., 2019. Automated identification of benthic epifauna with computer vision. Mar. Ecol. Prog. Ser. 615, 15—30. https://doi.org/10.3354/meps12925

Reeves, B.R., Dowty, P.R., Wyllie-Echeverria, S., Berry, H.D., 2007. Classifying the seagrass Zostera marina L. from underwater video: an assessment of sampling variation. J. Mar. Environ. Eng. 9 (1), 1—15.

Rimavičius, T., Gelžinis, A., Verikas, A., Vaičiukynas, E., Bačauskienė, M., Šaškov, A., 2018. Automatic benthic imagery recognition using a hierarchical two-stage approach. Signal Image Video P 12 (6), 1107—1114. https://doi.org/10.1007/s11760-018-1262-4

Rzhanov, Y., Mayer, L., 2004. Deep-sea image processing, 647—652, Oceans '04 MTS/IEEE Techno-Ocean '04 (IEEE Cat. No.04CH37600). https://doi.org/10.1109/OCEANS.2004.1405498

Šaškov, A., Dahlgren, T.G., Rzhanov, Y., Schläppy, M.L., 2015. Comparison of manual and semi-automatic underwater imagery analyses for monitoring of benthic hard-bottom organisms at offshore renewable energy installations. Hydrobiologia 756, 139—153. https://doi.org/10.1007/s10750-014-2072-5

Šiaulys, A., Vaičiukynas, E., Medelytė, S., Olenin, S., Šaškov, A., Buškus, K., Verikas, A., 2021. A fully-annotated imagery dataset of sublittoral benthic species in Svalbard. Arctic. Data Br. 35, 106823. https://doi.org/10.1016/j.dib.2021.106823

Smith Menandro, P., Cardoso Bastos, A., 2020. Seabed mapping: A brief history from meaningful words. Geosciences 10 (7), 273. https://doi.org/10.3390/geosciences10070273

Tan, M., Le, Q., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, 6105—6114. https://doi.org/10.48550/arXiv.1905.11946

Tomar, S., 2006. Converting video formats with FFmpeg. Linux J. 2006 (146), 10. https://www.linuxjournal.com/article/8517

Trygonis, V., Sini, M., 2012. photoQuad: a dedicated seabed image processing software, and a comparative error analysis of four photoquadrat methods. J. Exp. Mar. Biol. Ecol. 424 (425), 99—108. https://doi.org/10.1016/j.jembe.2012.04.018

Urra, J., Palomino, D., Lozano, P., González-García, E., Farias, C., Mateo-Ramírez, Á., Fernández-Salas, L.M., López-González, N., Vila, Y., Orejas, C., Puerta, P., 2021. Deep-sea habitat characterization using acoustic data and underwater imagery in Gazul mud volcano (Gulf of Cádiz, NE Atlantic). Deep-Sea Res. Pt. I 169, 103458. https://doi.org/10.1016/j.dsr.2020.103458

Vaičiūtė, D., Bucas, M., Bresciani, M., 2012. Validation of MERIS bio-optical products with *in situ* data in the turbid Lithuanian Baltic Sea coastal waters. J. Appl. Remote Sens. 6 (1), 063568. https://doi.org/10.1117/1.JRS.6.063568

Wentworth, C.K., 1922. A scale of grade and class terms for clastic sediments. J. Geol. 30 (5), 377—392. https://doi.org/10.1086/622910

Yakubovskiy, P., 2019. Segmentation Models. GitHub repository. URL: https://github.com/qubvel/segmentation_models (accessed on 2 August 2023).

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid Scene Parsing Network. Proc. CVPR IEEE 6230—6239. https://doi.org/10.1109/CVPR.2017.660