ISSN 1213-6972 (print)
ISSN 1213-6964 (online)
Journal of WSCG
http://www.wscg.eu
Vol.30, No-1-2, 2022

Exploring the necessity of mosaicking for underwater imagery semantic segmentation using deep learning

Kazimieras Buskus
Kaunas University of Technology, Kaunas, Lithuania
kazimieras.buskus@ktu.edu

Evaldas Vaiciukynas
Kaunas University of Technology, Kaunas, Lithuania
evaldas.vaiciukynas@ktu.lt

Saule Medelyte
Klaipeda university, Klaipeda, Lithuania
saule.medelyte@ku.lt

Andrius Siaulys
Klaipeda university, Klaipeda, Lithuania
andrius.siaulys@jmtc.ku.lt

ABSTRACT

Deep learning applications are attracting considerable interest nowadays and image analysis pipelines are no exception. Benthic studies often rely on the subjective evaluation of video material recorded using underwater drones. The demand for automatic image segmentation and quantitative evaluation arises due to the large volume of video data collected. This study performed a semantic segmentation task by training the PSPNet architecture with ResNet-34 backbone for 50 epochs using imagery prepared by simply extracting a few video frames or stitching a multitude of frames into a large 2D mosaic. Mosaicking is a particularly resource-intensive step, therefore, the possibility to skip such preprocessing would result in a more rapid analysis. The effect on the resulting segmentation quality was investigated by estimating the seabed coverage of three classes (*Furcellaria lumbricalis, Mytilus edulis trossulus*, and boulders) in a video material obtained from the Baltic Sea. Segmentation success, measured by intersection over union, varied between 0.56 and 0.84, usually slightly better for frames than for the mosaic overall. Absolute differences in estimated coverage were negligible (mosaic vs. frames): 0.24% vs. 1.26% for furcellaria, 0.44% vs. 2.46% for mytilus, and 4.02% vs. 2.06% for boulders. Due to the differences between predicted coverage and the mosaic-based ground truth being in an acceptable range, the findings suggest that the mosaicking step could be safely skipped in favor of a few equally spaced sample frames.

Keywords

underwater imagery, mosaicking, semantic segmentation, deep learning, PSPNet, ResNet, Baltic sea

1 INTRODUCTION

Maritime space is increasingly used for renewable energy installations, oil and gas exploitation, naval shipping and fishing, ecosystem monitoring and biodiversity conservation, aquaculture production, and many other purposes. The demand for maritime space requires integrated planning and management strategies focused on solid scientific knowledge and reliable seabed mapping [Men20], with underwater images [Urr21] being one of the most widely used seabed mapping input sources. The key advantage of underwater imagery is its simplicity, enabling the rapid collection of vast amounts of data, especially through the use of underwater drones and hence cost-effectiveness. Unfortunately, only a small part of the information stored in these image archives is being extracted due to labor-intensive and time-consuming analysis procedures. A promising way to process large amounts of images is computer-aided analysis, i.e., converting seabed video to 2D mosaic maps, image segmentation, and quantifying segmentation results. However, the mosaicking step is computationally demanding, and the resulting photomosaic often differs with respect to the tool used (see Fig. 1). A comprehensive list of various, primarily commercial, tools was evaluated for airborne imagery by [Son16] with Pix4DMapper found to be the most precise and Autostich the fastest.

Automatic segmentation of underwater imagery, compared to other types of images, is a new and challenging direction of research. According to a survey [Gra17] the first publications on 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 [Rim18]. New deep learning architectures of neural networks could replace image features and help analyze images more effectively, accurately

(a) CCOM [Rzh06] (b) LAPM v1 [Mar13] (c) Microsoft ICE 2 (d) ArcSoft PM 6 (e) Sample frames

Figure 1: Converting underwater video material (from Atlantic Ocean) to images: marine mosaics for benthic studies (a-b), commercial software mosaics for panoramic view (c-d), and sample of equally spaced frames (e).

and quickly than ever before. The initial efforts to apply deep learning to underwater images concerned corals [Alo19] and other broad categories [Liu20, Isl20] (such as fish, plants, divers, and stones) without preoccupation with the sea floor and, therefore, without the need to stitch images through mosaicking with the goal of reconstructing a precise floor map for the estimation of biologically valid organism counts or visual coverage.

The most similar related works that use deep learning and convolutional architectures to explicitly segment seabed images are recent techniques for estimating the coverage of seagrass [Mar18, Wei19, Bur20], macroalgae [Bal20], and kelp [Mah20]. The discrepancy between the expert-identified and model-predicted coverage in absolute percentage points was between 0.54% and 9.88% for kelp [Mah20] species (see Table 6 there), while the remaining mentioned seabed studies reported only segmentation accuracy without a comparison of the resulting coverage percentages.

This study evaluates both semantic segmentation success and the resulting coverage estimates using seabed imagery in the form of 2D mosaics, or several sample frames from short 30 s video transects as input to a convolutional architecture deep learning model. Both mosaics, as in Fig. 1 (a), and frames, as in Fig. 1 (e), were annotated by marine biologists to solve the detection task of 2 biological species – *Furcellaria lumbricalis* and *Mytilus edulis trossulus* – and 1 geological class – boulders. Performed experiments use mosaic-based annotations as a golden standard to measure how much coverage summaries differ when skipping a resource-intensive mosaicking step in favor of a few sample frames from an underwater video.

2 UNDERWATER IMAGERY

The underwater material was filmed in coastal and offshore reefs of Lithuanian marine waters of the south-eastern Baltic Sea. In the coastal area, the video was taken by scuba divers at 4–7 meters depth using a hand-held video camera with 1920×1080 resolution along multiple 10–meter transects (designations SM_07_1, SM_07_2, SM_08_1A, SM_08_1B). Additionally, in the offshore region at 35 – 40 meters depth, a remotely operated vehicle (ROV) equipped with a vertically mounted camera (3 CCD, high-quality Leica Dicomar lenses, and 10× optical zoom) with 1920×1080 resolution and a lighting system consisting of 16 LED

in 4 × 4 array was deployed, and two 30 second long transects (designations Denoflit_30s and Denoflit_2) were filmed. The raw video material was later transformed into 2D mosaics or several relevant frames were picked for each transect.

In this study, the video mosaicking method, developed by the Center for Coastal and Ocean Mapping [Rzh06], was used. The method consists of the following steps:

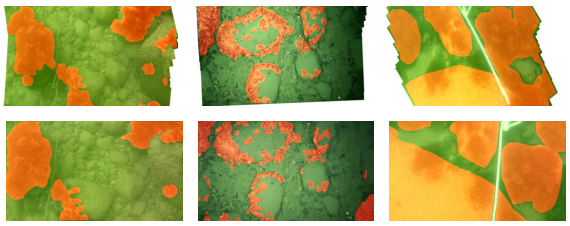
1. To reduce processing time, video transects of 30 s had frame rate reduced from 50 to 5 fps and frame size from 1920×1080 to 960×540.
2. The roll and pitch of the filming platform were adjusted by image transformations and some video enhancements were applied to each frame.
3. The enhanced footage was subjected to automatic frame-to-frame pair-wise registration (a method that calculates the overlap of neighboring frames).
4. Using the pair-wise registration data from the previous step, video mosaics were constructed from non-enhanced video.

For an alternative 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 tool *ffmpeg*. The aim of several representative sample frames was to cover the video material with the least overlap between frames to roughly correspond with the mosaic's visual scope. Summary of the prepared image data and classes represented in the images is in Table 1, where the size of the resulting 2D mosaics and the corresponding number of representative frames (of 960×540 resolution) can be compared.

The prepared imagery (large mosaics and many fixed-size frames) was annotated by two marine biologists (without overlap) using polygons and striving for pixel-level accuracy in an online collaborative annotation platform Labelbox [Rie20]. Examples of these annotations for the mosaic segment and the corresponding frame are shown in Fig. 2.

Transect	Mosaic size	Frames	Classes
SM07_1	2434×8774	16	*Furcellaria* Boulder
SM07_2	5021×5107	12	*Furcellaria*
SM08_1A	4191×5379	12	*Furcellaria* Boulder
SM08_1B	4745×5379	12	*Furcellaria* Boulder
Denoflit_2	2434×8774	11	*Mytilus* Boulder
Denoflit_30s	1580×5480	11	*Mytilus*

Table 1: Summary of underwater imagery used.



(a) *Furcellaria* (b) *Mytilus* (c) Boulders

Figure 2: Mosaic (*top*) and frame (*bottom*) annotations for (roughly) the same transect region.

In general, all visible objects of 3 classes were annotated: biological *Furcellaria* species had 102 instances in 3 mosaics and 120 instances in frames; biological *Mytilus* species had 148 instances in 2 mosaics and 62 in frames; geological Boulder class had 167 instances in 3 mosaics and 166 in frames.

3 METHODS

The prepared underwater imagery, either in the form of large mosaics or representative frames, was patched using a sliding window idea to prepare training and testing data for the deep learning model with convolutional architecture. Training patches were augmented to increase data amount and as a simple form of regularization. The evaluation was performed by splitting the transects in half to achieve a 2-fold cross-validation. Segmentation success was measured by the intersection over union metric and by comparing visual coverage estimates.

3.1 Image patching and augmentation

Due to the limitations of the available computational resources, both mosaics and frames were sliced into overlapping 288×288 size patches. The overlap was the result of sliding window or block processing idea with vertical and horizontal strides of 144 pixels (see Fig. 3). The mosaics contained many white pixels, as a result of the mosaicking process, so only patches with a minimum of 70% non-white pixels were considered as input images. Furthermore, 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) [Cha10] – were used on input image patches.

3.2 Convolutional architecture

In the experiments, we used a deep convolutional neural network with pyramid spatial pooling architecture - PSPNet [Zha17] model with ImageNet pre-trained ResNet-34 [He16] as the backbone. The model was implemented using the *Keras* framework (version 2.3.1),

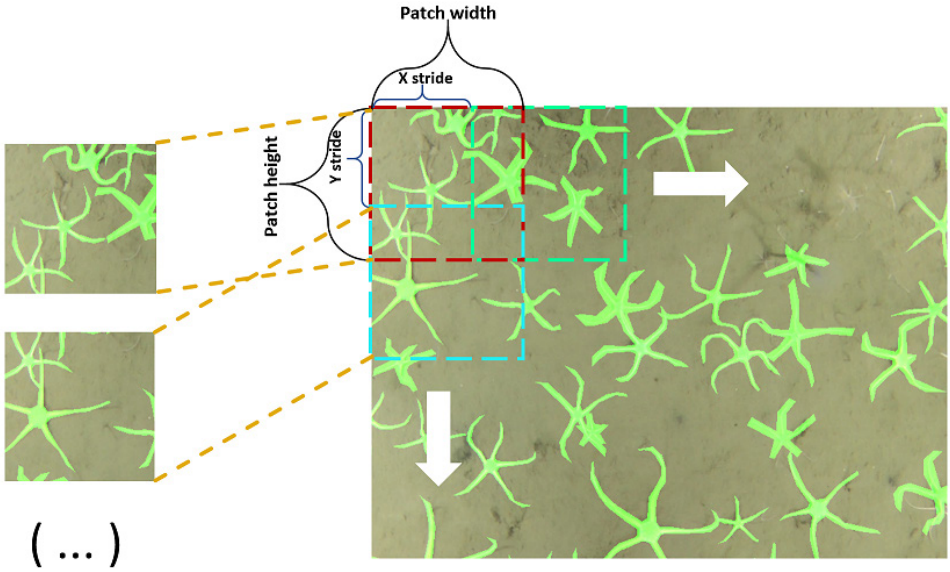


Figure 3: Block processing of underwater photos into overlapping patches of 288×288 sized input images.

running on the *Tensorflow* back-end (version 2.1.0), with the help of the package *segmentation-models* (version 1.0.1) [Yak19]. Model training settings were as follows: batch size (number of images used for the weight tuning step) - 8 images, training duration - 50 epochs, downsampling rate (backbone depth in the PSPNet model) - 8, minimized loss - additive combination of Jaccard [Jac08] and Focal [Lin17] losses. Final model had 2 746 058 parameters (2 737 860 trainable).

3.3 Evaluation of segmentation

Evaluation of segmentation performance was done using 2-fold transect-stratified cross-validation (CV), where each transect was split in half – top and bottom parts – and training was performed on one part while testing on the other part. For example, after training on all bottom parts of mosaics (first half of the corresponding frame set), testing was performed on all top parts and vice versa. Both training and testing were carried out using 288×288 size images and after obtaining results on all mosaic (or frames) patches tested, overlapping parts were averaged and a class threshold of 0.5 was used to obtain the final predicted mask. Then the intersection over union (IOU) – a commonly used measure of segmentation success, comparing the ground truth with the predicted mask – was used to summarize the results of both testing folds of 2-fold CV in a micro-average fashion. In addition, final prediction masks were used to estimate the visual coverage of the class in question. The coverage itself was interpreted as a ratio between predicted or the ground truth masks with either relevant pixels (excluding white pixels) in the mosaic setting and all pixels in the frame setting, see the Pixels column in Tables 2– 4.

4 EXPERIMENTS

We used 2D mosaics and representative transect frames for training and testing the convolutional neural network model for two biological and one geological classes.

4.1 Setup

The hardware used was as follows: Intel(R) Core(TM) i7-8700 CPU @3.2 GHz, 32 GB of operating memory, NVIDIA RTX 2070 with 8 GB of memory. Software used: Windows 10 Enterprise (build 1809) 64-bit operating system, CUDA 10.1, CuDNN 6.4.7 and Python 3.6.8. During training, 4608 patches were used for *Furcellaria* class mosaic setting, 4200 for frame settings; *Mytilus* class - 1944 for mosaic, 1848 for frames; Boulder - 9810 mosaic, 8652 patches in frame settings.

4.2 Results

Segmentation performance. Segmentation performance was summarized using the intersection over union (IOU) metric. The amount of imagery used in experiments is reported as Pixels column in Tables 2 – 4, with frames usually having slightly fewer pixels overall, except for *Furcellaria* class transects. Results demonstrate that the best achieved IOU score was 84% using mosaics for *Furcellaria* class segmentation (see Table 2). The IOU results between mosaics and frames indicated slight differences in all but one case for *Furcellaria* class, where the absolute difference in IOU scores was almost 10%. Also, in two out of four transects for *Furcellaria* class, the frames had a better IOU score (by 1.04% and 1.5% points). For *Mytilus* class (see Table 3), frames had marginally better IOU results (by 2.8% and 6.1% points). However, mosaics appeared to be more advantageous for the geological

Furcellaria transects	Mosaic		Frames		
	Pixels (mln.)	IOU	Pixels (mln.)	IOU	Δ IOU
SM07_1_bio	9.04	0.703	8.29	0.718	-0.015
SM07_2_bio	7.06	0.839	6.22	0.793	0.046
SM08_1A_bio	6.64	0.661	6.22	0.765	-0.104
SM08_1B_bio	6.85	0.726	6.22	0.726	0
Totals:	29.59	0.711	26.96	0.746	-0.035

Table 2: Segmentation performance, as measured by the IOU score, for *Furcellaria* class using mosaics and frames.

Mytilus transects	Mosaic		Frames		
	Pixels (mln.)	IOU	Pixels (mln.)	IOU	Δ IOU
Denoflit_2_bio	6.56	0.560	5.70	0.621	-0.061
Denoflit_30s_bio	5.17	0.671	5.70	0.699	-0.028
Totals:	11.73	0.613	11.40	0.6649	-0.051

Table 3: Segmentation performance, as measured by the IOU score, for *Mytilus* class using mosaics and frames.

Boulder transects	Mosaic		Frames		
	Pixels (mln.)	IOU	Pixels (mln.)	IOU	Δ IOU
Denoflit_2_geo	6.56	0.661	5.70	0.592	0.070
SM07_1_geo	9.04	0.606	8.29	0.603	0.003
SM08_1A_geo	6.63	0.819	6.22	0.798	0.021
SM08_1B_geo	6.85	0.802	6.22	0.808	-0.006
Totals:	29.09	0.744	26.44	0.733	0.011

Table 4: Segmentation performance, as measured by IOU, for Boulder class using mosaics and frames.

boulder class (see Table 4), with a maximum difference of 7% points. Overall, total IOU scores were higher for frames than mosaics by 3.5% for *Furcellaria* and 5.1% for *Mytilus* class (see Totals in Tables 2 and 3), except for the boulder class (see Totals in Table 7) with 1% point difference.

Segmentation totals. When summarizing the segmentation performance by the Totals in Tables 2–4 we can make the following insights. Segmentation success, measured by the IOU metric, was in an acceptable range between 61.3% (for *Mytilus* mosaics) and 74.6% (for *Furcellaria* frames). Segmentation performance using frames was better for biological classes (by 3.5% for *Furcellaria* and 5.1% for *Mytilus* transects), but slightly worse for the geological boulder class (by 1.1%).

Coverage estimates. With respect to visual coverage estimates, it is important to note that for all but the *Mytilus* class, transects have very different coverage levels: for example, transect *SM07_1_bio* had 9.75% while transect *SM08_1B_bio* had 57.48% coverage for the *Furcellaria* class (see the Mosaic GT column in Table 5). Three deltas summarize the results of visual coverage estimates in Tables 5–7, measuring the absolute difference in coverage from the mosaic ground truth: the first delta (Mosaic Δ DL) shows the effect of using predictions from mosaics, the second delta (Frames Δ GT) shows the effect of using annotated frames instead of mosaics without any prediction, and the third delta (Frames Δ DL) shows the effect of using predictions from frames. Biological classes had negligible differences when using mosaic predictions with an underestimate of 0.86% and an overestimate of 0.41%, while geological class had greater differences with an underestimate of 1.61% and an overestimate of 6.12%. Surprisingly, even annotators were unable to achieve a high correspondence in visual coverage estimates with slight differences for geological class and more considerable differences for biological classes, even reaching an underestimate of 8.24% points for *Furcellaria* transect with the highest coverage (transect *SM07_2_bio*). The last and most important differences in our study were observed using frames predictions with an underestimate of 5.89% and an overestimate of 8.41%, both for the *Furcellaria* class. The differences for other classes were distributed relatively uniformly in that range. There was a tendency to overestimate geological class and underestimate biological classes when estimating visual coverage from frame predictions.

Coverage totals. When summarizing the coverage estimates by the Totals in Tables 5–7 we can make the following insights. Taking mosaics as the ground truth,

Furcellaria transects	Mosaic			Frames			
	GT	DL	Δ DL	GT	Δ GT	DL	Δ DL
SM07_1_bio	9.75	9.44	0.31	8.32	1.43	9.01	0.74
SM07_2_bio	11.91	11.05	0.86	9.88	2.03	9.25	2.66
SM08_1A_bio	43.15	43.83	-0.68	47.56	-4.41	49.04	-5.89
SM08_1B_bio	57.48	57.07	0.41	49.23	8.24	49.07	8.41
Totals:	28.81	28.57	0.24	27.18	1.63	27.55	1.26

Table 5: Coverage estimates for *Furcellaria* class. Abbreviations: GT – results from ground truth annotations; DL – results from deep learning model-based predictions; Δ – difference from mosaic-wise ground truth annotations.

Mytilus transects	Mosaic			Frames			
	GT	DL	Δ DL	GT	Δ GT	DL	Δ DL
Denoflit_2_bio	18.11	17.92	0.20	15.25	2.86	16.25	1.86
Denoflit_30s_bio	22.83	22.07	0.76	21.61	1.21	19.21	3.62
Totals:	20.19	19.75	0.44	18.43	1.76	17.73	2.46

Table 6: Coverage estimates for *Mytilus* class. Abbreviations: GT – results from ground truth annotations; DL – results from deep learning model-based predictions; Δ – difference from mosaic-wise ground truth annotations.

Boulder transects	Mosaic			Frames			
	GT	DL	Δ DL	GT	Δ GT	DL	Δ DL
Denoflit_2_geo	29.65	28.05	1.61	30.62	-0.97	25.84	3.82
SM07_1_geo	34.95	41.07	-6.12	34.52	0.43	36.24	-1.29
SM08_1A_geo	76.41	81.95	-5.54	77.05	-0.64	81.17	-4.77
SM08_1B_geo	76.55	81.70	-5.15	76.72	-0.17	80.90	-4.35
Totals:	53.01	57.03	-4.02	53.61	-0.60	55.08	-2.06

Table 7: Coverage estimates for boulder class. Abbreviations: GT – results from ground truth annotations; DL – results from deep learning model-based predictions; Δ – difference from mosaic-wise ground truth annotations.

we can notice lower coverage for biological classes (28.81% for *Furcellaria* and 20.19% for *Mytilus*) than for the geological class (53.01% for boulders). Expert-based annotations of frames deviated from the annotations of mosaics only slightly: more for biological classes (by 1.63% for *Furcellaria* and by 1.76% for *Mytilus*) and less for the geological class (by 0.6% for boulders). Model-based predictions when training on the transects' bottom half and testing on the top half (or vice versa), if compared to previously mentioned deviations of expert-based frames annotations, deviated only slightly more with underestimates of 1.26% for *Furcellaria* and 2.46% for *Mytillus* classes and an overestimate of 2.06% for boulders class.

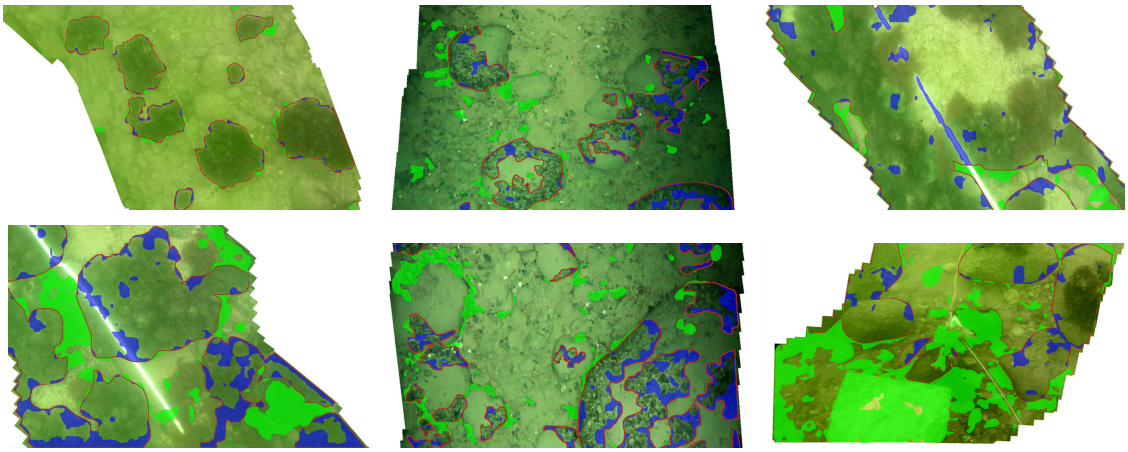
Visual comparison. Examples of segmentation predictions are provided in Figures 4–5, where we can compare how poor results differ from acceptable with segmentation false positives/negatives in green/blue colors.

5 CONCLUSIONS

Segmentation success was better than average with intersection over union varying between 0.56 and 0.84, depending on the class and transect, but slightly better for frames than for the mosaics overall. Lower visual coverage estimates from ground truth mosaic annotations were for biological classes (28.81% for *Furcellaria* and 20.19% for *Mytilus*) than for the geological class (53.01% for boulders).

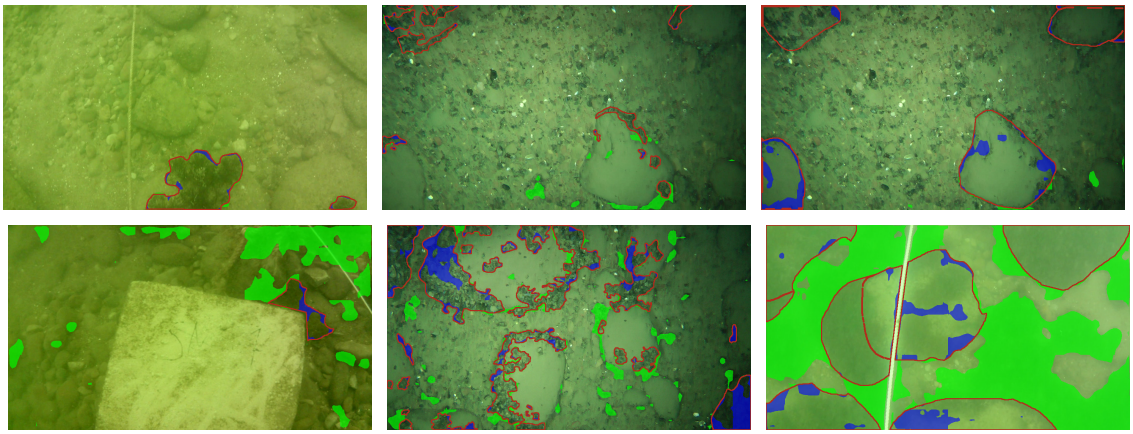
Expert-based annotations of frames deviated from the annotations of mosaics only slightly (largest deviation of 1.76%), model-based predictions - a bit more (largest deviation of 2.46%). The largest differences were observed for *Mytilus* class, which was composed of many tiny objects and had the lowest coverage. Due to the deviations being in an acceptable range, the reported results of visual coverage estimates suggest that the mosaicking step could be safely skipped in favour of a few equally spaced sample frames.

The main limitation of the experiments performed is the balance between training and testing data amounts, and the transect-stratified type of cross-validation, ensuring that the model has seen part of the tested transect during training. Future work should address these limitations.



(a) *Furcellaria* (b) *Mytilus* (c) Boulders

Figure 4: Segmentation success using mosaics: acceptable (*top*) and poor (*bottom*) results. Color coding: false negative pixels are marked in blue, false positive pixels in green, and ground truth annotations are outlined in red.



(a) *Furcellaria* (b) *Mytilus* (c) Boulders

Figure 5: Segmentation success using frames: acceptable (*top*) and poor (*bottom*) results. Color coding: false negative pixels are marked in blue, false positive pixels in green, and ground truth annotations are outlined in red.

6 ACKNOWLEDGMENTS

This work 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). The authors thank the Labelbox [Rie20] team.

REFERENCES

[Alo19] Alonso, I., Yuval, M., Eyal, G., Treibitz, T., and Murillo, A. C. CoralSeg: Learning coral segmentation from sparse annotations. Journal of Field Robotics 36, No. 8, 2019, pp. 1456–1477. doi:10.1002/rob.21915.

[Bal20] Balado, J., Olabarria, C., Martínez-Sánchez, J., Rodríguez-Pérez, J. R., and Pedro, A. Semantic segmentation of major macroalgae in coastal environments using high-resolution ground imagery and deep learning. International Journal of Remote Sensing 42, No. 5, 2020, pp. 1785–1800. doi:10.1080/01431161.2020.1842543.

[Bur20] Burguera, A. Segmentation through patch classification: A neural network approach to detect Posidonia oceanica in underwater images. Ecological Informatics 56, 2020, p. 101053. doi:10.1016/j.ecoinf.2020.101053.

[Cha10] Chao, L. and Wang, M. Removal of water scattering. In: 2010 2nd International Conference on Computer Engineering and Technology (ICCET). 2, 2010, pp. 2–35. doi:10.1109/ICCET.2010.5485339.

[Gra17] Gracias, N., Garcia, R., Campos, R., Hurtos, N., Prados, R., Shihavuddin, A., Nicosevici, T., Elibol, A., Neumann, L., and Escartin, J. Application Challenges of Underwater Vision. In: Computer Vision

in Vehicle Technology. 2017, pp. 133–160. doi:10.1002/9781118868065.ch7.

[He16] He, K., Zhang, X., Ren, S., and Sun, J. Deep Residual Learning for Image Recognition. In: 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.

[Isl20] Islam, M. J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., Enan, S. S., and Sattar, J. Semantic Segmentation of Underwater Imagery: Dataset and Benchmark. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2020. doi:10.1109/iros45743.2020.9340821.

[Jac08] Jaccard, P. Nouvelles Recherches Sur la Distribution Florale. Bulletin de la Societe Vaudoise des Sciences Naturelles 44, 1908, pp. 223–70. doi:10.5169/seals-268384.

[Lin17] Lin, T., Goyal, P., Girshick, R., He, K., and DollÃ¡r, P. Focal Loss for Dense Object Detection. In: 2017 IEEE International Conference on Computer Vision (ICCV). 2017, pp. 2999–3007. doi:10.1109/ICCV.2017.324.

[Liu20] Liu, F. and Fang, M. Semantic Segmentation of Underwater Images Based on Improved Deeplab. Journal of Marine Science and Engineering 8, No. 3, 2020, p. 188. doi:10.3390/jmse8030188.

[Mah20] Mahmood, A., Ospina, A. G., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., Fisher, R. B., and Kendrick, G. A. Automatic Hierarchical Classification of Kelps Using Deep Residual Features. Sensors 20, No. 2, 2020, p. 447. doi:10.3390/s20020447.

[Mar13] Marcon, Y., Sahling, H., and Bohrmann, G. LAPM: a tool for underwater large-area photo-mosaicking. Geoscientific Instrumentation, Methods and Data Systems 2, No. 2, 2013, pp. 189–198. doi:10.5194/gi-2-189-2013.

[Mar18] Martin-Abadal, M., Guerrero-Font, E., Bonin-Font, F., and Gonzalez-Cid, Y. Deep Semantic Segmentation in an AUV for Online Posidonia Oceanica Meadows Identification. IEEE Access 6, 2018, pp. 60956–60967. doi:10.1109/access.2018.2875412.

[Men20] Menandro, P. S. and Bastos, A. C. Seabed Mapping: A Brief History from Meaningful Words. Geosciences 10, No. 7, 2020, p. 273. doi:10.3390/geosciences10070273.

[Rie20] Rieger, B., Rasmuson, D., and Sharma, M. Labelbox: the leading training data platform for data labelling. 2020. URL: http://labelbox.com.

[Rim18] Rimavičius, T., Gelžinis, A., Verikas, A., Vaičiukynas, E., Bačauskienė, M., and Šaškov, A. Automatic benthic imagery recognition using a hierarchical two-stage approach. Signal, Image and Video Processing 12, No. 6, 2018, pp. 1107–1114. doi:10.1007/s11760-018-1262-4.

[Rzh06] Rzhanov, Y., Mayer, L., Beaulieu, S., Shank, T., Soule, S., and Fornari, D. Deep-sea Geo-referenced Video Mosaics. In: OCEANS 2006. 2006. doi:10.1109/oceans.2006.307018.

[Son16] Song, H., Yang, C., Zhang, J., Hoffmann, W. C., He, D., and Thomasson, J. A. Comparison of mosaicking techniques for airborne images from consumer-grade cameras. Journal of Applied Remote Sensing 10, No. 1, 2016, p. 016030. doi:10.1117/1.jrs.10.016030.

[Urr21] 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., Rivera, J., Henry, L.-A., and Rueda, J. L. Deep-sea habitat characterization using acoustic data and underwater imagery in Gazul mud volcano (Gulf of Cádiz, NE Atlantic). Deep Sea Research Part I: Oceanographic Research Papers 169, 2021, p. 103458. doi:10.1016/j.dsr.2020.103458.

[Wei19] Weidmann, F., Jager, J., Reus, G., Schultz, S. T., Kruschel, C., Wolff, V., and Fricke-Neuderth, K. A Closer Look at Seagrass Meadows: Semantic Segmentation for Visual Coverage Estimation. In: OCEANS 2019 - Marseille. 2019. doi:10.1109/oceanse.2019.8867064.

[Yak19] Yakubovskiy, P. Segmentation Models. GitHub repository, 2019. 2019. URL: https://github.com/qubvel/segmentation_models.

[Zha17] Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. Pyramid Scene Parsing Network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017, pp. 6230–6239. doi:10.1109/CVPR.2017.660.