Semantic Segmentation in Underwater Ship Inspections: Benchmark and Data Set

Maryna Waszak, *Member, IEEE*, Alexandre Cardaillac, Brian Elvesæter, Frode Rødølen, and Martin Ludvigsen, *Member, IEEE*

***Abstract*—In this article, we present the first large-scale data set for underwater ship lifecycle inspection, analysis and condition information (LIACI). It contains 1893 images with pixel annotations for ten object categories: defects, corrosion, paint peel, marine growth, sea chest gratings, overboard valves, propeller, anodes, bilge keel and ship hull. The images have been collected during underwater ship inspections and annotated by human domain experts. We also present a benchmark evaluation of state-of-the-art semantic segmentation approaches based on standard performance metrics. Consequently, we propose to use U-Net with a MobileNetV2 backbone for the segmentation task due to its balanced tradeoff between performance and computational efficiency, which is essential if used for real-time evaluation. Also, we demonstrate its benefits for in-water inspections by providing quantitative evaluations of the inspection findings. With a variety of use cases, the proposed segmentation pipeline and the LIACI data set create new promising opportunities for future research in underwater ship inspections.**

***Index Terms*—Data set, semantic segmentation, supervised machine learning, underwater inspection.**

I. Introduction

ANNOTATED data sets of underwater ship hull inspections for semantic segmentation are scarce. In this section, we present our motivation for creating such a publicly available data set by describing how in-water ship inspections are conducted and how semantic segmentation would make the process more efficient.

The rest of this article is organized as follows. Section II describes the collection of data and the creation of the data set used for the training of the selected semantic segmentation models. Section III presents and discusses the experimental results of the benchmark evaluation. Section IV points out future directions in improving the data set and how a semantic segmentation model could aid other research topics in underwater computer vision. Finally, Section V concludes this article.

A. Underwater Ship Inspections

Visual inspections are rigorously applied in different domains of our lives. With increasing exploitation of marine resources, significant attention is being drawn to the importance of underwater ship inspections. As of today, the monitoring and inspection of marine vessels is performed based on recurrent visual observations and assessments of structural condition either in dry-dock or underwater. The main purpose of these inspections is to assist with the examination of the external coating, as well as detection of corrosion or marine growth. Inspections in dry-dock are significantly costlier than in-water inspections in addition to longer downtime of the ship. Therefore, ship hull inspections performed underwater are increasing in popularity. With the technological advances in the field of autonomous underwater vehicles, the need for automated data processing becomes inevitable as the manual reviewing and processing of collected videos, images, and other nondestructive inspection data (e.g., ultrasonic thickness measurements) becomes unfeasible [1].

B. Semantic Segmentation

The advances in computer vision provide ways for increasing reliability and effectiveness for acquiring, managing, integrating, and interpreting the acquired inspection data at a minimum cost while reducing the need for tedious and often unreliable data analysis by a human expert. Specifically, automated processing of image and video data is a great source of quantitative insight that can complement the largely qualitative information obtained from conventional visual inspections. In contrast to land images, however, the underwater environment poses several challenges for automated image processing. The images may be deteriorated by different artifacts, such as water turbidity, floating particles, severe absorption, reflections, scattering of light, nonuniform illumination, various noises, low contrast and monotonous colors. See Fig. 1 for some examples of mentioned artifacts.

This work focuses on semantic image segmentation in the domain of underwater ship inspections and how it can aid the inspection procedure by providing additional insight from the acquired underwater video data. Semantic segmentation refers to pixelwise classification, a class label is assigned to each pixel

Manuscript received 10 March 2022; revised 25 August 2022; accepted 26 October 2022. This work was supported in part by The Research Council of Norway through LIACI Project under Grant 317854, and in part by European Union's Horizon 2020 Research and Innovation Program through BugWright2 Project under Grant 871260. *(Alexandre Cardaillac and Maryna Waszak contributed equally to this work.) (Corresponding author: Maryna Waszak.)*

Associate Editor: B. Thornton.

Maryna Waszak and Brian Elvesæter are with SINTEF AS, 0373 Oslo, Norway (e-mail: maryna.waszak@sintef.no; brian.elvesater@sintef.no).

Alexandre Cardaillac and Martin Ludvigsen are with the Department of Marine Technology, Norwegian University of Science and Technology, 7034 Trondheim, Norway (e-mail: alexandre.cardaillac@ntnu.no; martin.ludvigsen@ntnu.no).

Frode Rødølen is with VUVI AS, 5035 Bergen, Norway (e-mail: frode@vuvi.no).

The data set is made publicly available for noncommercial use on https://liaci.sintef.cloud.

Digital Object Identifier 10.1109/JOE.2022.3219129

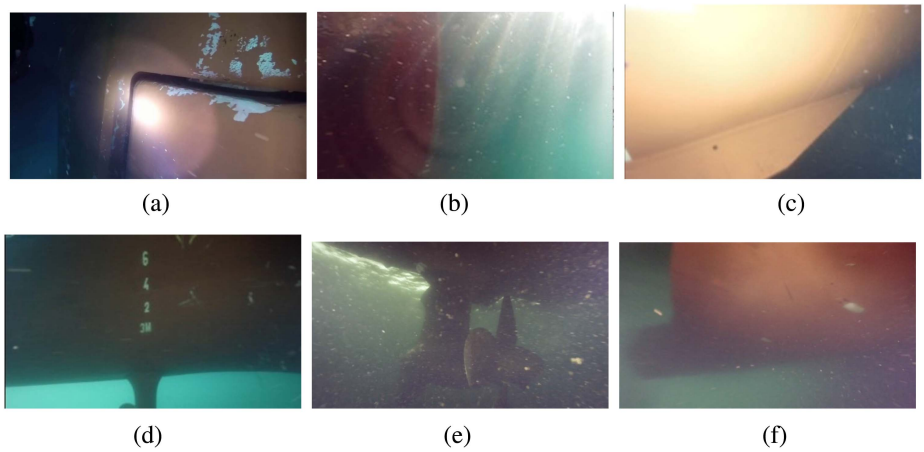

Fig. 1. Common artifacts in underwater imagery. (a) Light beam. (b) Light scattering. (c) Reflections. (d) Scratches on lens. (e) Floating particles. (f) Water turbidity.

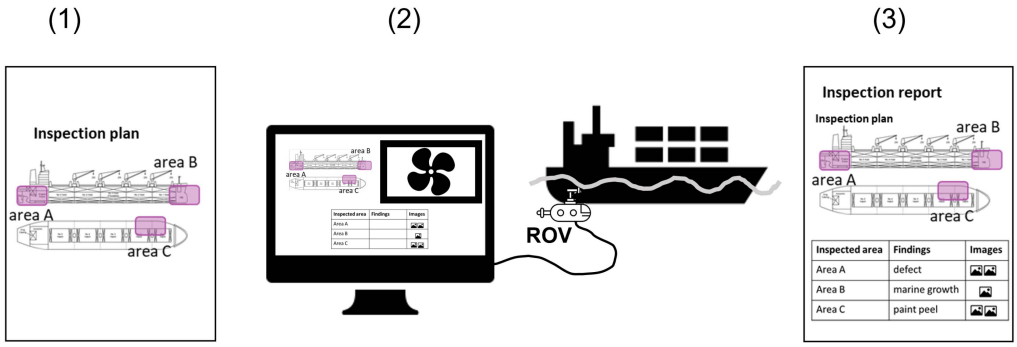

Fig. 2. Current inspection workflow is performed in three separate steps: 1) planning, 2) data acquisition, and 3) report creation.

of the image. It is a well-studied problem as it is a key for scene understanding. It decomposes the scene into objects or categories, which are significant semantic regions. Recent methods involving a deep learning approach have achieved outstanding results [2], [3], [4]. The current state-of-the-art segmentation networks have been mostly proposed for and applied to medical image analysis, driver-less cars or other surface applications. It remains to be shown that those successful networks can be successfully applied to underwater segmentation tasks. In this work, we aim at closing this gap through a benchmark evaluation on our data set.

C. Available Data Sets

Currently, manually labeled data sets such as ImageNet [5], ADE20K [6], PASCAL [7], and COCO [8] play a significant role in improving machine vision tasks and driving research in new directions. Data sets with underwater imagery such as SUIM [9] or Seagrass [10] exist that aim at the semantic segmentation task or the classification of fish [11] or marine growth [12] species. Although works related to the detection and segmentation of relevant classes and objects in the domain of visual surface inspections as marine growth, corrosion, and cracks exist, the underlying data sets remain undisclosed or are inaccessible [13], [14], [15], [16], [17], [18], [19].

We wanted to create a publicly available data set that aims at the task of semantic segmentation of underwater ship inspection images. This data set is meant to be used as a starting point for underwater scene understanding and improved machine vision in the domain of in-water ship inspections.

D. Lifecycle Inspection, Analysis and Condition Information (LIACI) Use Case

Here, we worked with a combination of commercial hardware and software for conducting underwater ship inspections. Several experts were involved and the inspections were performed in different steps as depicted in Fig. 2: 1) planning, 2) data acquisition, and 3) report creation. The software and hardware involved is named the LIACI system. The introduced use case is from two Norwegian companies: VUVI AS,[1] which is a commercial provider of underwater ship hull inspections, and Posciom AS,[2] which is the provider of the video tagging and management platform Seekuence.

The current data acquisition setup consists of the underwater remotely operated vehicle (ROV) and two separate video streams. One stream is used for the navigation of the ROV and

[1][Online]. Available: vuvi.no
[2][Online]. Available: www.posicom.no

(a) (b) (c) (d)

Fig. 3. ROVs that were used for data collection during underwater ship inspections. (a) JM Robotics BlueROV2. (b) JM Robotics JM HD1. (c) VideoRay Pro 4. (d) Blueye Pioneer.

TABLE I
NUMBER OF IMAGES COLLECTED BY DIFFERENT ROV TYPES

ROV name	Image resolution	Image count	Number of different ships
VideoRay Pro 4	640 × 480	284	5
JM Robotics BlueROV2	1920 × 1080	725	4
JM Robotics JM HD1	1280 × 720	837	7
Blueye Pioneer	1920 × 1080	47	1

the second one for video frame annotation, where interesting frames in the video are marked to be further evaluated. The ROVs are supplied by different commercial vendors (see Section II-A on data collection for further details). After the video data are acquired, the annotated video snippets are screened in a postprocessing step and snap shots are extracted for a final inspection report. The acquired and annotated data are archived for later reference.

The current workflow is tedious and time consuming and does not incorporate any automated data processing. We propose to use semantic segmentation to identify and quantify different metrics relevant for the ship inspection procedure. By automatic annotation and quantitative analysis of video data, the inspection report can be created without human interaction. Hence, the effort in the third step in the current workflow can be significantly reduced.

II. LIACI DATA SET

This section describes the collection of inspection video data and the image extraction process. It illustrates the difficulties specific to labeling underwater inspection data and explains the classes that were chosen as labels for the annotation task. Further, it shows how the images were annotated, and the resulting statistical properties of the images. We also included an evaluation on the similarity of the images in the data set.

A. Data Collection

Videos from 16 underwater ship inspections were collected by the commercial inspection provider VUVI AS using two different ROVs from JM Robotics AS[3] and one ROV from VideoRay[4] with an in-built filter from LYYN.[5] The names of the vessels remain secret due to nondisclosure agreements with the ship owners. Additionally, at the research vessel Gunnerus,[6] one video of the hull was acquired with the Pioneer drone from Blueye.[7] Fig. 3 shows the drones that were used for data collection, and in Table I, the individual image count that was chosen for the data set. The videos were recorded at different locations in the Norwegian Sea off the Norwegian coast. From these videos, a representative collection of images was extracted by the ROV operator during the video recording and in the postprocessing step preparing the inspection report.

Imaging tasks in an underwater environment are challenging. Even though some of characteristics are generalizable, many are dependent on the location and its situation. The underwater visibility is mainly affected by the penetration of the light and the water turbidity. Because of this, it is important that the proposed data set presents image diversity in terms of underwater scene conditions. This makes it possible to reduce the classwise water condition specific overfitting when training a model. Even though the data were acquired only at the Norwegian coast, we can observe a range of visibility conditions. The variety of ships presents different feature combinations, which is an important aspect to further improve the robustness.

The images were extracted by the ship inspector during video recording and in the postprocessing step to reflect the status of the inspected areas. These images were used to train an image classifier to find images in similar classes to ramp up the image count. In the first sweep, the inspector usually extracted approximately 50–100 images from the video. We trained a vision transformer multilabel classifier with Microsoft Custom

[3][Online]. Available: www.jmrobotics.no

[4][Online]. Available: videoray.com
[5][Online]. Available: www.lyyn.com
[6][Online]. Available: www.ntnu.edu/gunnerus
[7][Online]. Available: www.blueye.no

TABLE II
OVERVIEW OF ANNOTATED CLASSES WITH ASSOCIATED DESCRIPTION AND MASK COLOR

Group	Class	Description	Mask color
Ship parts	Ship hull	The main ship structure.	Blue
	Propeller	All revolving structures on the ship.	Purple
	Bilge keel	A stabilizing structure on the ship hull to reduce rolling motion.	Orange
	Anode	Sacrificial anodes that provide galvanic cathodic protection of submerged metal structures from corrosion.	Cyan
	Sea chest grating	Sea chests are intake reservoirs for water piping systems on a ship. They are protected by removable gratings.	White
	Overboard valve	Usually located on the sides of the ship. They are round openings on the ship hull that serve as in and outlets.	Turquoise
Inspection criteria	Corrosion	Oxidized metal parts of the ship.	Yellow
	Paint peel	Any damage to the condition of the anti-fouling coverage on the ship hull. That is coating, paint, or other surface treatment that is used on a ship to control or prevent attachment of unwanted marine organisms.	Red
	Marine growth	The accumulation of aquatic organisms such as micro-organisms, algae, and animals on surfaces and structures immersed in or exposed to the aquatic environment. Bio-fouling types can include soft bio-fouling and hard calcareous bio-fouling.	Green
	Defect	All other defects that are neither corrosion, marine growth, nor paint peel.	Pink

Vision[8] [20] and indexed the videos to find images for the classes of interest. This way we could ramp up the image count to a total of 1893 images and also mimic the inspectors' choice for the data from the research vessel Gunnerus where no inspector was involved. Fig. 11 summarizes the steps visually.

B. Data Labeling

A total of ten different labels divided into two categories were proposed. They were selected to provide relevant and detailed information that could be used for an automated or aided inspection. The first category corresponds to the physical parts of a ship that can be found underwater, while the second category is about what can be found on the surface of the ship that is not originally part of it. The latter category is called inspection criteria because it corresponds to what the inspector is looking for when performing an inspection. These are often subject to evolve over time, e.g., disappear after maintenance, change over time, and reappear again. We have often noncanonical viewpoints and only some iconic images, thus we focus mostly on categories with clear boundaries. However, due to natural water turbidity that increases with the distance from the camera, the ship hull and other relatively big ship parts do not have clear boundaries.

An overview of the classes is given in Table II with a description for each class. The colors are used to differentiate the labels in the processed scenes. In the majority of cases, these two categories overlap each other, providing information about the location of the inspection criteria.

The selected classes cover a large part of the image while minimizing the "blank" part of the image, i.e., without annotation. These parts frequently correspond to the underwater background.

We created guidelines for labeling to have consensus among the annotators to mitigate some of the annotation difficulties. Specifically, it is not a trivial task separating marine growth, paint peel, and corrosion, as can be seen in Fig. 4. Here, it is extremely difficult to separate the different classes as they usually appear overlapping each other and rarely on their own.

The annotation task was performed by two annotators using the Microsoft Azure Machine Learning Studio[9] web-based platform. The annotation method consisted of layered polygons that when combined should cover the entire underwater scene without the background. After completion of the data set, all the images and associated annotations were reviewed again and corrected where necessary by the same annotators to guarantee

[8][Online]. Available: https://www.customvision.ai/

[9][Online]. Available: https://ml.azure.com

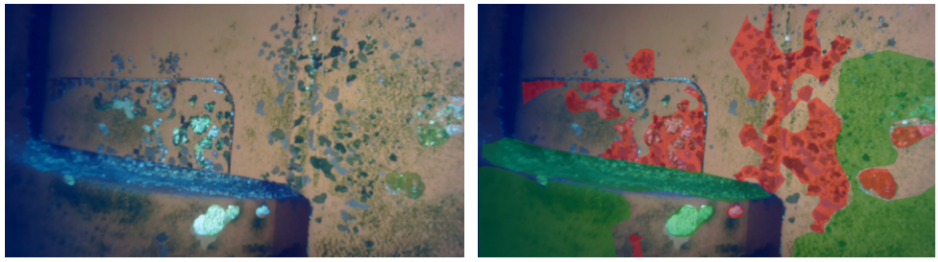

Fig. 4. Example of an image where separating the labels for paint peel and marine growth is challenging due to overlaps. The raw image is shown on the left and the same image with overlapped segmentation results on the right.

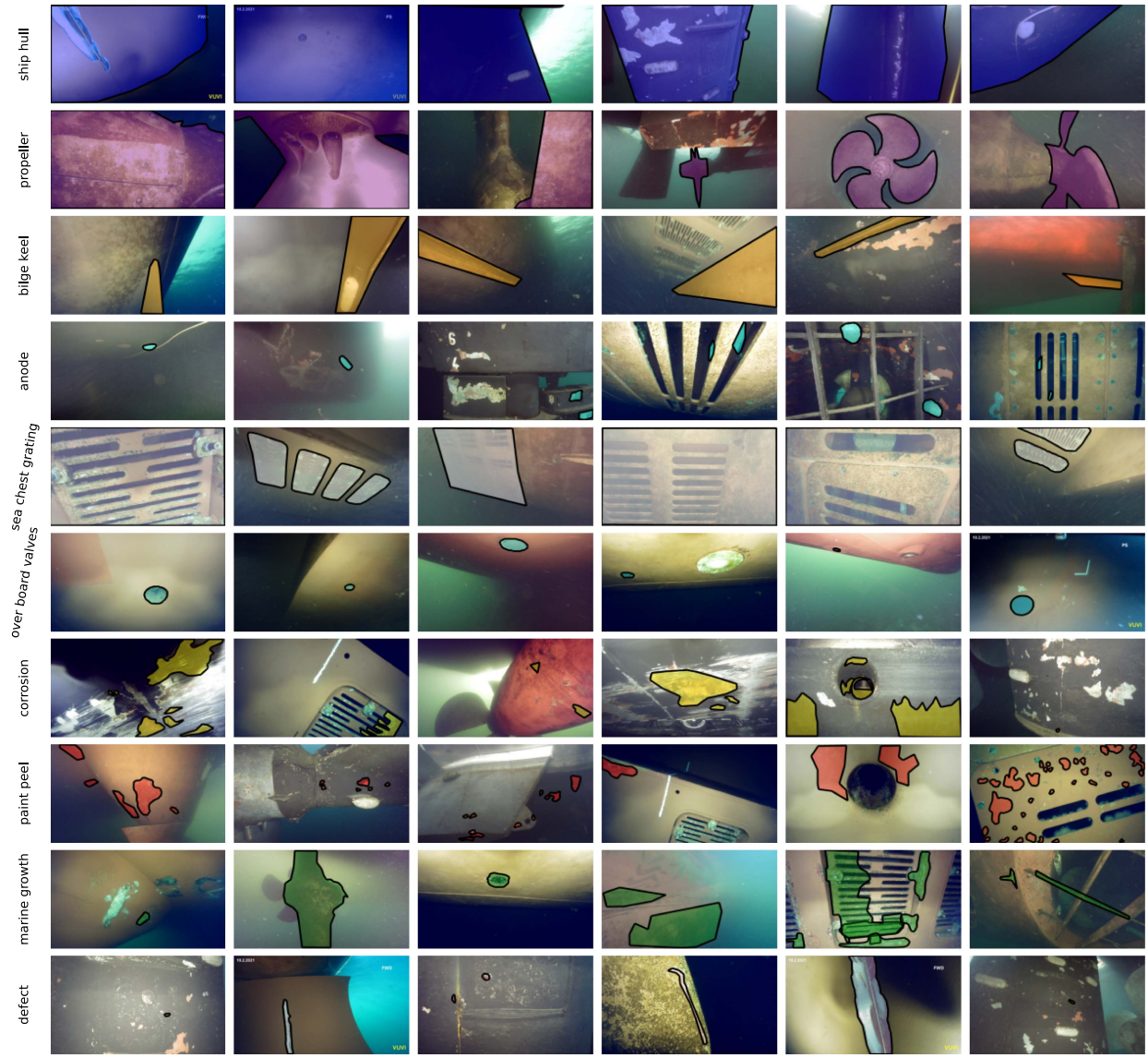

Fig. 5. Samples of annotated images in the LIACI data set for each class.

high fidelity annotations. The samples of annotated images for each class are depicted in Fig. 5, with one class per row.

A representative sample of the data set consisting of 100 images was sent to a professional ship inspector to assess the quality and precision of the annotations. The inspector had access to the labeling tools, allowing him to update the masks based on his knowledge. We used his review as ground truth to compute the precision, recall, and F1 score for each class and to determine if there was any action to be taken. The results are given in Table III for each class and metrics. The "ship parts" category is very accurate, this was expected since all the subparts are very easily recognizable and can hardly be confused. For

TABLE III
ANNOTATORS LABEL EVALUATION WITH TWO SCORE METRICS, PRECISION, AND RECALL FOR ALL CLASS CATEGORIES, INCLUDING WHERE MARINE GROWTH AND PAINT PEEL WERE CONSIDERED AS ONE SINGLE CLASS IN THE "COMBINED" COLUMN

Metric	Ship hull	Propeller	Bilge keel	Anode	Sea chest grating	Overboard valve	**Mean**
Precision	98.41	99.51	98.85	99.00	98.46	99.74	99.00
Recall	98.67	99.98	99.93	99.00	99.98	99.00	99.42
F1 Score	98.54	99.74	99.39	99.00	99.21	99.37	99.21
Metric	Defect	Corrosion	Paint peel	Marine growth	**Mean**		**Combined**
Precision	99.00	89.99	92.68	80.94	90.65		95.66
Recall	99.00	99.06	86.04	92.00	94.02		97.26
F1 Score	99.00	94.31	89.23	86.11	92.30		96.45

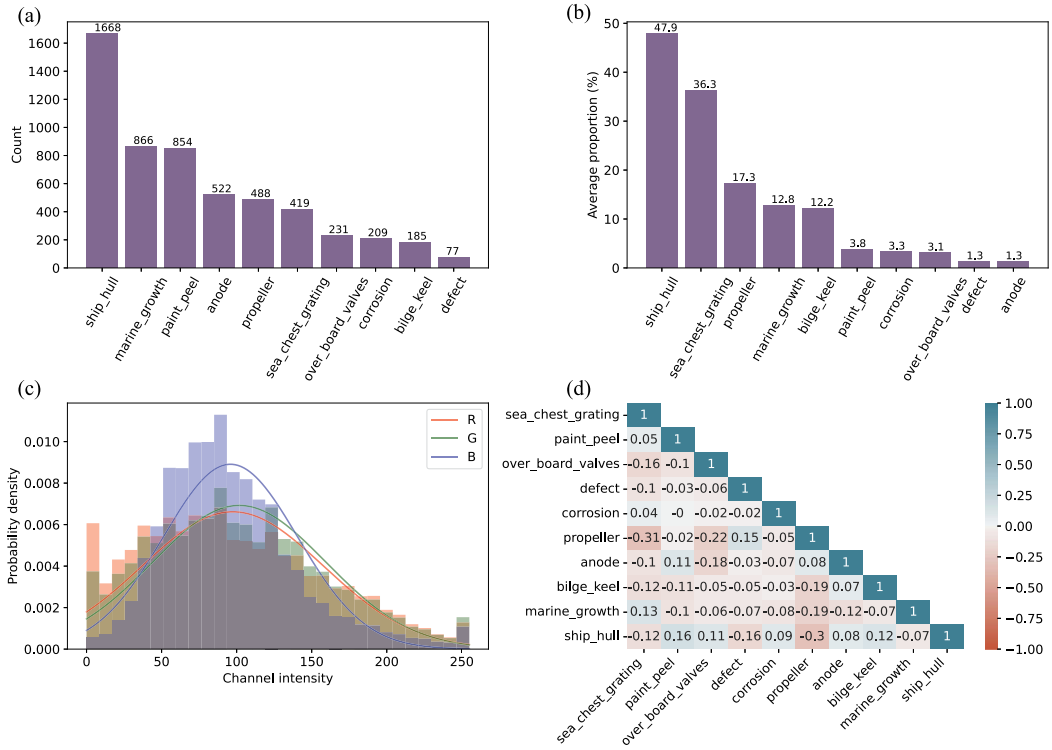

Fig. 6. Overview over the image and label statistics. (a) Number of annotated images per class. (b) Average proportion of annotated pixels in images for each class. (c) Distribution of the pixel intensities in each channel. (d) Pairwise correlation of the labels occurrences in the data set based on the Pearson method.

the "inspection criteria," however, the distinction between the subparts is not as easy, especially with marine growth and paint peel, and sometimes corrosion. Based on the inspector's review, some were misclassified, but overall, the three metrics remain acceptable and indicate the quality of the data set labels.

C. Data Set Presentation

The proposed data set contains 1893 RGB images alongside their pixelwise annotations for semantic segmentation. Images with different aspect ratio and resolution are included, e.g., 1920×1080, 1280×720 and 640×480. Detailed statistics of the images and labels are shown in Fig. 6. Since pixel intensity value is the primary information stored within pixels, it is the most popular and important feature used in computer vision. The intensity value for each pixel consists of three values for the color images. In the presented data set, we observed that the blue channel is over-represented compared to the red and green colors that is easily explained by the underwater domain where the images were collected.

The pairwise correlations of the labels are calculated using Pearson's correlation coefficient r. It quantifies the linear relationship between two distributions based on the covariance and standard deviation

$$r_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{1}$$

with the two distributions X and Y. r ranges from -1, the perfect negative correlation, to $+1$, the perfect positive correlation. Therefore, since the correlation matrix presented shows a good

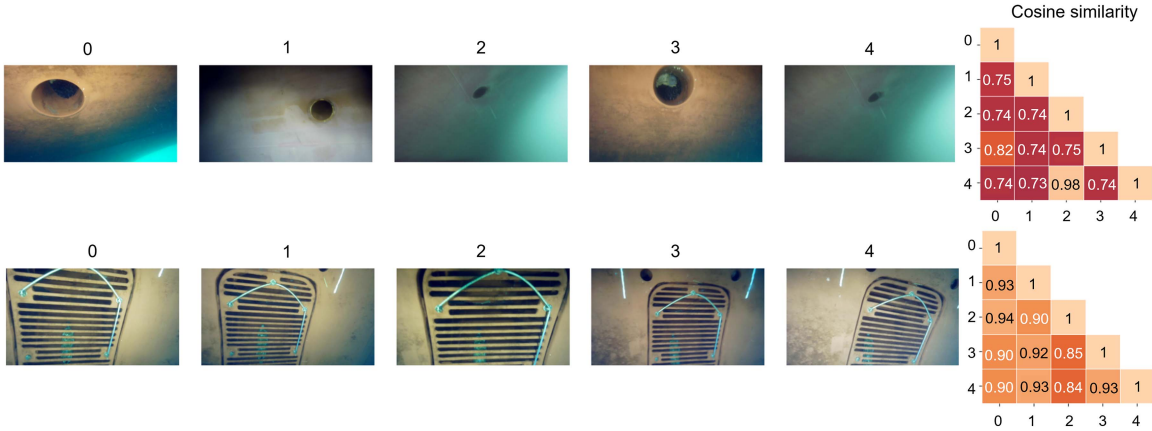

Fig. 7. Similar images of a sea chest grating and an overboard valve with corresponding pairwise Cosine similarity index.

diversification of classes in images, there is no single combination that makes it possible to find a class based on another. There is then a good distribution/representation of classes in the data set imagewise. There are no strong correlations but some still exist, for example, with the pair propeller/sea chest grating, which is negatively correlated with a value of -0.31. We could think it should be stronger since sea chest gratings are never present on propellers, but some images in the data set contain both at different locations because of the viewpoint of the ROV.

We extracted the images from videos. Therefore, we had grounds to assume that similar images might be among the images in the data set. To quantitatively evaluate how many similar images there are, we calculated a feature vector by extracting the last fully connected layer from the ResNet101 classifier pretrained with ImageNet as provided by PyTorch.[10] We chose ResNet101 as recommended in [21] and an initial naive evaluation provided good results. The calculated image vectors were then used to calculate pairwise Cosine similarity, where an index of 1 means that images are exactly the same and 0 a complete orthogonality. The similarity index follows a normal distribution with a mean and standard deviation of 0.64 ± 0.07, indicating that we have similar images in our data set since the closer the values are to 1, the higher the similarity. Fig. 7 shows example similar images of a sea chest grating and an overboard valve. For different cut-off values for the Cosine similarity measure, Fig. 8 shows the number of unique images in the data set. If we were to choose a cut-off at 0.90 and consider the same labels are present, the data set will still have 1561 images left. Thus, this is the value we recommend to filter out images that are too similar as also confirmed by a qualitative visual evaluation.

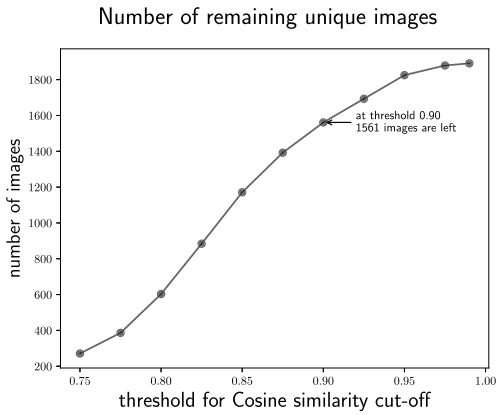

Fig. 8. Number of remaining images after filtering at different thresholds of the Cosine similarity metric and same classes being present on the image.

III. Benchmark Evaluation

This section describes the motivation behind the chosen segmentation models for the benchmarking evaluation and presents the results of the evaluation in detail. It is done using multiple combinations of encoders and decoders to prove the capability of the data set to be used for training and converge at a reasonable rate.

A. Semantic Segmentation Models

For the benchmark evaluation, multiple state-of-the-art deep convolutional neural network (CNN) models were considered. Often, CNN models can be divided into two parts: an encoder and a decoder. The way the layers are arranged in the encoder network corresponds to the architectural element called backbone. For example, a model such as MobileNetV2 can be used as an encoder for the UNet model, which retains the decoding layers [22]. During the evaluation, backbones based on other models were often included instead of vanilla CNNs; these were pretrained on ImageNet [5]. Also, some segmentation models were utilized multiple times but with different backbones. The complete list of models is displayed in Table IV.

All the models were implemented in Python using the TensorFlow libraries [30]. The same hardware setup is used for all models: NTNU IDUN computing cluster [31], with an NVIDIA Tesla P100 GPU for training, and a laptop with an NVIDIA Geforce GTX 1060 for testing. For the training, the data set was augmented by applying random image transformations

[10][Online]. Available: https://pypi.org/project/img2vec-pytorch/

TABLE IV
LIST OF SEGMENTATION MODELS AND THEIR BACKBONES USED FOR THE BENCHMARK ALONG WITH THEIR NUMBER OF PARAMETERS, INPUT RESOLUTION, AND AVERAGE INFERENCE FRAME RATE AS COMPUTED ON A SINGLE NVIDIA GTX 1060 GPU

Referred as	Segmentation model	Backbone	Number of parameters	Resolution	FPS
UNet+MobileNetV2	UNet [23]	MobileNetV2 [24]	8,048 M	320 × 256	23.17
UNet+VGG	UNet [23]	VGG16 [25]	23,753 M	320 × 256	17.86
DeepLabV3	DeepLabV3 [26]	Vanilla	41,256 M	320 × 320	10.91
FPN+MobileNetV2	FPN [27]	MobileNetV2 [24]	5,220 M	320 × 256	17.16
FPN+ResNet50	FPN [27]	ResNet50 [28]	26,922 M	320 × 256	14.71
PSP+MobileNetV2	PSP [29]	MobileNetV2 [24]	1,662 M	336 × 336	25.68
SegNet+ResNet50	SegNet [3]	ResNet50 [28]	15,014 M	320 × 256	18.44
SuimNet+RSB	SuimNet [9]	RSB [9]	3,866 M	320 × 240	17.96
SuimNet+VGG	SuimNet [9]	VGG16 [25]	12,228 M	320 × 256	16.01

from a defined list. They consisted of rotation, shear and zoom effects, as well as, horizontal flip and slight brightness shift. This augmentation was done in addition to image removal based on the similarity measure. This might have made the models less accurate but able to generalize better.

After filtering the data set based on the similarity metric presented in Section II-C with a threshold of 0.90, we divided the remaining 1561 images into a training subset composed of 1370 (87.8%) images and a testing subset with 191 (12.2%) images. These numbers are the result of ensuring a uniform distribution of classes in the training and testing subsets.

B. Evaluation Criteria

To measure the performance of the models, multiple criteria were considered. To evaluate the correctness of the pixelwise classification, two supervised evaluation methods were utilized: the Intersection over Union (IoU) and the F1 Score. The former, also known as the Jaccard Index, is one of the most used metrics for semantic segmentation tasks. It consists of the area of overlap between the predicted masks and the ground truth divided by the area of union between the prediction and the ground truth

$$\begin{aligned} \text{IoU} &= \frac{\text{Area of overlap}}{\text{Area of union}} \\ &= \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive} + \text{False Negative}}. \end{aligned} \quad (2)$$

It is also regarded as a region similarity metric.

The latter is also called the dice coefficient and provides the contour accuracy $\mathcal{F}1$

$$\mathcal{F}1 = \frac{2 \times \mathcal{P} \times \mathcal{R}}{\mathcal{P} + \mathcal{R}}. \quad (3)$$

It is defined as the harmonic mean of the precision $\mathcal{P}$ and recall $\mathcal{R}$ of the model.

Also, for the considered applications, time constraints are present. Therefore, the inference time needs to be taken into account. For real-time capabilities, a minimum of ten frames per second (FPS) are required. Also, because the segmentation task needs to be performed during data acquisition, it needs to be possible to run it on the operator laptop which might contain a low-cost GPU or sometimes rely just on the CPU.

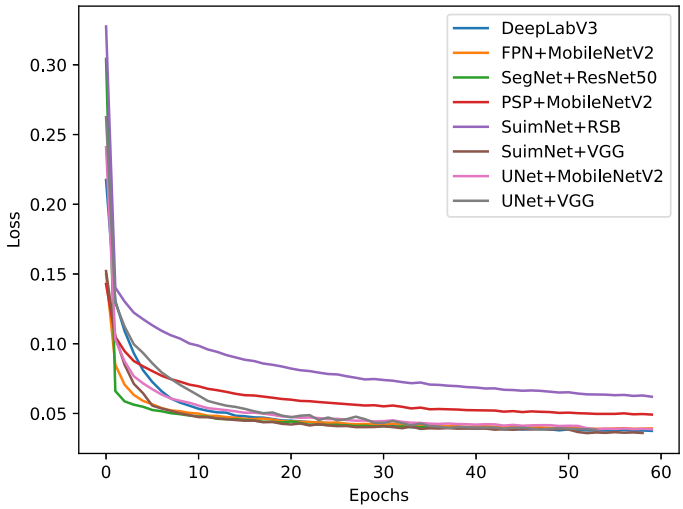

Fig. 9. Training loss over epochs of the considered models until epoch 60.

C. Quantitative and Qualitative Analysis

A benchmark evaluation with state-of-art deep learning segmentation models showed that good results can be obtained with all of the selected models. Also, all models show similar convergence behavior, with SegNet converging fastest, as depicted in Fig. 9. Table V lists the results of the benchmark evaluation. SegNet with the ResNet50 backbone provides the best results for the class of ship parts with a mean IoU of 86.07 and a mean F1 score of 88.17. The inference time for the PSP model with the MobileNetV2 backbone showed the best time of 25.68 FPS. Over all models, the segmentation accuracy for marine growth and paint peel is not as good compared to other classes. Several reasons could be the cause of this effect. Annotating the classes of marine growth and paints peel is challenging due to high variability of shapes and structures. Also, these label classes tend to naturally overlap as marine growth usually starts growing in areas with paint defects where the antifouling coating is missing. Corrosion also usually appeared on areas with paint

TABLE V
BENCHMARK FOR SEMANTIC SEGMENTATION WITH TWO SCORE METRICS F1 SCORE AND IOU FOR ALL CLASS CATEGORIES

Metric	**Model**	Ship hull	Propeller	Bilge keel	Anode	Sea chest grating	Overboard valve	**Mean**	Defect	Corrosion	Paint peel	Marine growth	**Mean**	**Combined**
IoU	UNet+MobileNetV2	77.52	83.82	89.13	82.2	90.03	93.06	85.96	92.76	83.36	47.64	61.37	71.28	80.09
	UNet+VGG16	80.81	84.68	88.21	78.75	90.58	91.44	85.74	93.79	81.16	45.98	58.32	69.81	79.37
	PSPNet+MobileNetV2	60.65	67.97	87.38	74.28	86.61	89.74	77.77	95.97	87.98	51.23	54.84	72.50	75.66
	DeeplabV3	74.10	79.04	87.92	74.53	80.72	92.44	81.46	94.82	79.60	37.38	52.20	66.00	75.28
	FPN+MobileNetV2	58.95	73.83	79.22	71.38	65.01	90.06	73.08	95.13	86.61	43.58	45.18	67.63	70.90
	FPN+ResNet50	55.54	65.27	76.07	74.25	48.82	88.09	68.01	95.93	87.76	43.53	47.77	68.74	68.30
	SuimNet+RSB	69.76	66.32	81.61	68.26	78.14	88.51	75.43	92.22	81.75	36.67	45.47	64.03	70.87
	SuimNet+VGG	72.38	73.09	77.57	79.94	83.85	89.61	79.41	94.20	83.56	38.56	54.65	67.74	74.74
	SegNet+ResNet50	76.85	85.46	87.75	80.40	93.80	92.17	86.07	95.83	87.03	43.38	63.62	72.46	80.63
F1 Score	UNet+MobileNetV2	82.26	85.56	90.18	84.25	91.29	93.84	87.90	93.00	84.19	51.82	66.07	73.77	82.25
	UNet+VGG16	85.22	86.04	89.06	81.04	91.95	92.28	87.60	94.04	81.72	50.50	62.71	72.24	81.46
	PSPNet+MobileNetV2	69.86	70.71	88.02	75.70	88.80	90.53	80.6	96.07	88.18	53.16	57.61	73.75	77.86
	DeeplabV3	79.12	81.19	88.78	76.36	82.30	93.14	83.48	95.07	80.06	41.77	55.64	68.13	77.34
	FPN+MobileNetV2	68.07	76.04	79.93	73.39	67.04	91.14	75.93	95.34	86.79	46.02	47.85	69.00	73.16
	FPN+ResNet50	65.52	67.12	77.40	76.11	50.68	88.98	70.97	96.01	87.94	46.19	50.31	70.11	70.63
	SuimNet+RSB	76.19	69.10	82.44	70.62	79.86	89.26	77.91	92.28	81.82	40.10	49.60	65.95	73.13
	SuimNet+VGG	77.98	75.07	78.81	82.08	85.21	90.40	81.59	94.45	84.22	42.90	59.51	70.27	77.06
	SegNet+ResNet50	81.79	87.3	88.88	82.84	95.33	92.89	88.17	96.09	87.58	47.39	67.78	74.71	82.79

peel but is less difficult to label resulting in better prediction results. Therefore, we performed another model training and evaluation round where the classes of marine growth and paint peel were merged. The results show that the accuracy of the merged label class could be increased for all models by almost ten points.

Another observation was that small objects disappeared due to downsampling of the images to the model resolution. Therefore, labels are reduced to only few pixels such that some models are no longer able to detect such areas, e.g., marine growth and paint peel. Dark areas on ship hulls in overboard valves, as well as, ship hull areas that were further away from the camera are not correctly identified by the models. Such qualitative observations are depicted in Fig. 10.

IV. FUTURE WORK

Image quality plays a major role in the performance of computer vision algorithms. Hence, seeking to improve image quality retrospectively would improve the results of automated image processing as suggested in [32]. The ULTIR data set [33] or the UIEB [34] could also be used as a starting point for identifying a method to prospectively guide the data acquisition to collect only images with sufficient quality. Image enhancement can be used as a preprocessing task before using it in the model. However, it is computationally expensive and not necessary to reach satisfying results, and hence was not included in this work but remains important for generalization purposes and more robust results. For these reasons, it will be considered in future work.

Our data set was solely collected off the Norwegian Sea. The visibility in waters differs significantly depending on the geographical location and light conditions. Hence, we believe that the data set would benefit from including videos from various waters.

There are several promising directions for future annotations on our data set. We currently only label few ship parts but this could be extended to other parts as the manoeuvring thruster, rudder or box cooler. Also, quantitative evaluation of potential defects inside the vessel water cooling system, which contains the impressed current antifouling anodes and should be monitored closely, could be a target for automated image processing algorithms. Further classes for defects (dents, cracks, rope around parts, scratches, etc.), paint peel (adhesion, blistering, cracking, cold flow, delamination, polishing-off, grounding), and marine growth (soft corals, sponges, hydroids, anemones, algae, tunicates, barnacles, mussels, tube worms, bryozoan, oysters, etc.) could be included in the annotations to follow the guidelines from the International Chamber of Shipping and The Baltic and International Maritime Council [35].

To improve the segmentation results further, the model could account for class correlation, i.e., overboard valves and anodes can only be on a ship hull and not on a propeller. Sea chest grating has to be surrounded by ship hull as well as overboard valves. Introducing such additional constraints would potentially improve model performance and reduce classification errors.

Future work will focus on propagating the segmentation masks onto the whole video to achieve thorough video indexing and to possibly aid algorithms for the calculation of structure from motion, simultaneous localization and mapping, and subsequent 3-D reconstruction of the inspected structures from the video data [36]. Here, the feature extraction step would benefit from adapting its calculation to the segmented object and use different features (e.g., ORB, SIFT) for objects with different semantic and visual properties.

V. CONCLUSION

Semantic understanding of videos in in-water ship inspections is critically important to facilitate quantitative analysis of collected image and video data. The existing solutions are application- and domain-specific as dedicated to the medical domain or autonomous driving for terrestrial vehicles and drones or industrial surface inspections in manufacturing. In this work, we attempt to address these limitations by presenting the first large-scale annotated data set for semantic segmentation of underwater ship inspection images. We described and made available a new

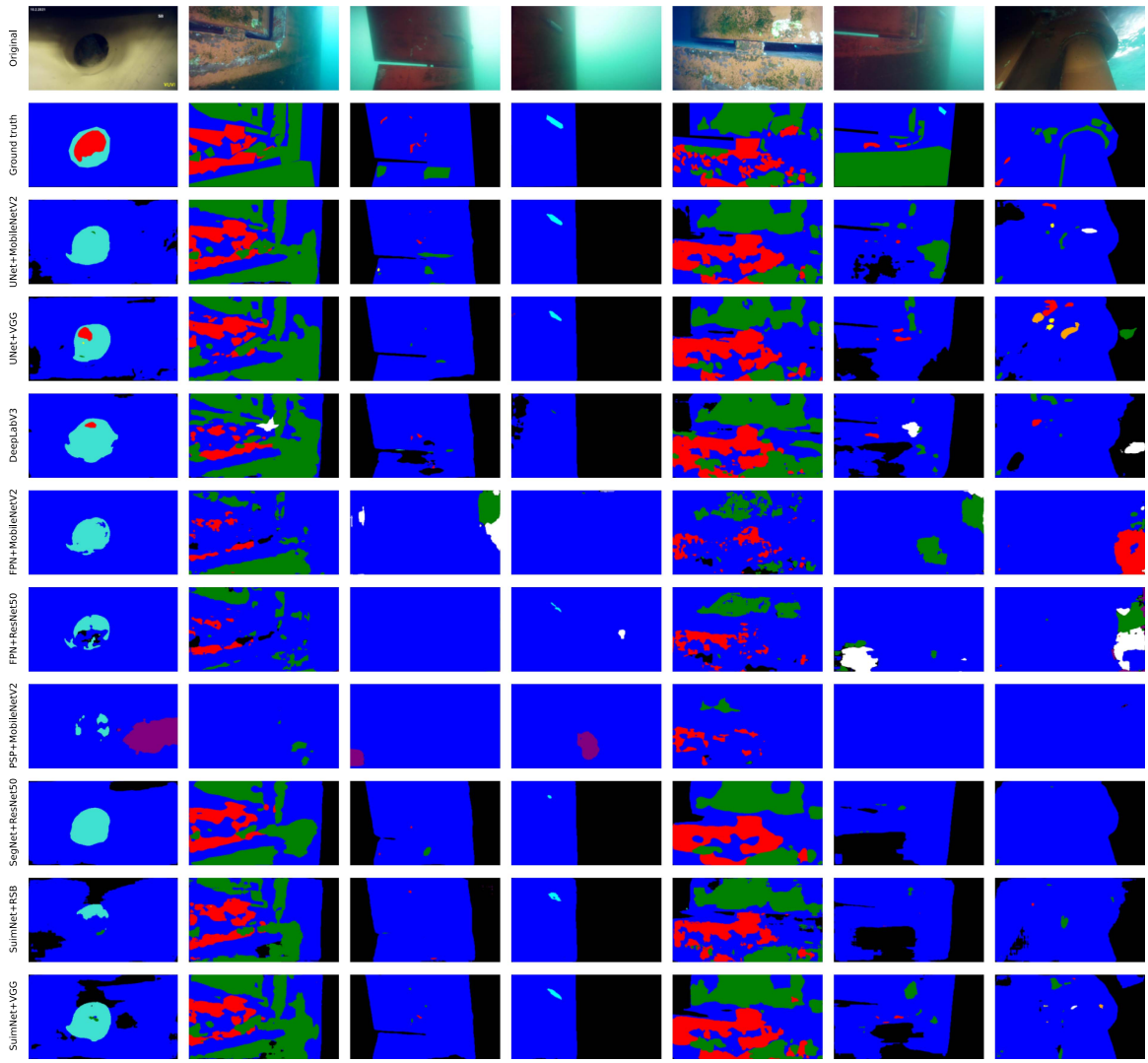

Fig. 10. Qualitative segmentation results for selected classes and models.

data set for detecting and segmenting objects in the domain of visual underwater ship inspections with ROVs. Involving two annotators and a reviewer, a collection of category instances was gathered, annotated, and organized to drive the advancement of object detection and segmentation algorithms. The proposed LIACI data set contains 1893 images with pixel annotations for ten object categories. The benchmark evaluation showed that the UNet segmentation model with the MobileNetV2 backbone provides the best overall performance in terms of segmentation results and inference time making it a good candidate for further investigations. Also, its architecture makes it possible to run the model on a consumer laptop without a GPU with an acceptable frame rate of up to 12 FPS. This is twice the frame rate that can be achieved with the SegNet model with ResNet50 backbone that has a frame rate of 5 FPS on average.

In comparison to humans, it is harder for segmentation models to extrapolate shapes, e.g., ship hull in the shade. The segmentation boundary of the target is not clear enough, the contour is incomplete, and the feature information is insufficient. Here, the annotations would benefit from other data sources as for example sonar or stereovision cameras. Also, having a 3-D model of the vessel could help to estimate and extrapolate the shape of the seen object. Therefore, enhancing the images with additional data from other sources would improve not only the quality of the annotations but also the model training process.

We also have to conclude that it is very difficult to annotate marine growth, paint peel, and corrosion separately. These classes often appear together and overlap. Therefore, we propose to fuse those classes and run unsupervised segmentation algorithms in a postprocessing step for further refinement.

Also, we deliberately did not exclude blurry images as we would like the data set to reflect the natural quality differences that appear during the data collection. We extracted the images from inspection videos which provided a natural augmentation of the data by providing different views of the objects. For example, the illumination conditions and the water turbidity were naturally changing when the ROV was capturing the object from different distances and angles.

The data set is made available for noncommercial use on https://liaci.sintef.cloud.

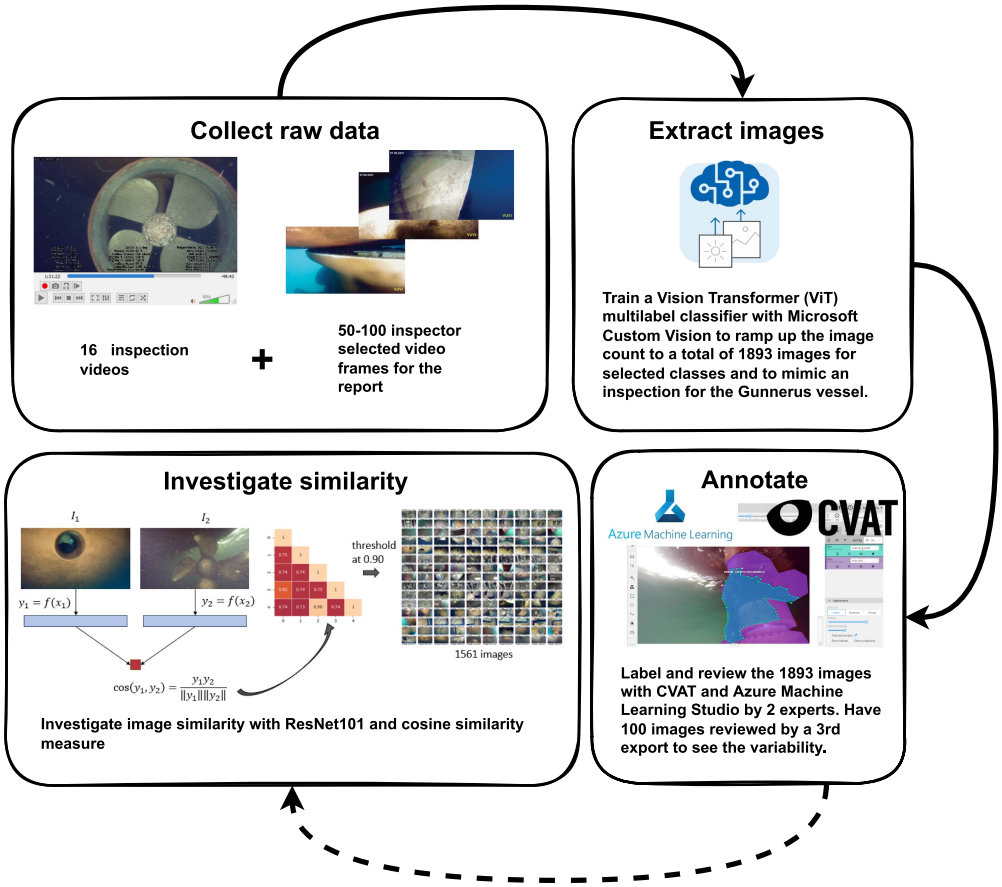

Fig. 11. Steps that were performed to create the data set.

APPENDIX

Fig. 11 shows the steps that were performed to create the data set.

ACKNOWLEDGMENT

The authors would like to thank A. Mohammed for valuable discussions and suggestions. A CC BY or equivalent license is applied to any Author Accepted Manuscript version arising from this submission, in accordance with the grant's open access conditions.

REFERENCES

[1] B. Ghosh, M. O'Byrne, F. Schoefs, and V. Pakrashi, *Image Based Damage Assessment for Underwater Inspections*. Boca Raton, FL, USA: CRC, Jan. 2019.

[2] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," *Appl. Soft Comput.*, vol. 70, pp. 41–65, 2018.

[3] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Los Alamitos, CA, USA, 2015, pp. 3431–3440.

[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.

[6] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ADE20 k dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5122–5130.

[7] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.

[8] T.-Y. Lin et al., "Microsoft coco: Common objects in context," in *Computer Vision–ECCV*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 740–755.

[9] M. J. Islam et al., "Semantic segmentation of underwater imagery: Dataset and benchmark," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 1769–1776.

[10] G. Reus et al., "Looking for seagrass: Deep learning for visual coverage estimation," in *Proc. IEEE/MTS OCEANS Conf., - Kobe Techno-Oceans*, 2018, pp. 1–6.

[11] D. Rathi, S. Jain, and S. Indu, "Underwater fish species classification using convolutional neural network and deep learning," in *Proc. 9th Int. Conf. Adv. Pattern Recognit.*, 2017, pp. 1–6.

[12] C. S. Chin, J. Si, A. Clare, and M. Ma, "Intelligent image recognition system for marine fouling using softmax transfer learning and deep convolutional neural networks," *Complexity*, vol. 2017, 2017, Art. no. 5730419.

[13] M. O'Byrne, V. Pakrashi, F. Schoefs, and a. B. Ghosh, "Semantic segmentation of underwater imagery using deep networks trained on synthetic imagery," *J. Mar. Sci. Eng.*, vol. 6, no. 3, Aug. 2018, Art. no. 93.

[14] S. K. Fondevik, A. Stahl, A. A. Transeth, and O. O. Knudsen, "Image segmentation of corrosion damages in industrial inspections," in *Proc. IEEE 32nd Int. Conf. Tools With Artif. Intell.*, 2020, pp. 787–792.

[15] F. Bonnín-Pascual and A. Ortiz, "Detection of cracks and corrosion for automated vessels visual inspection," in *Proc. Int. Conf. Catalan Assoc. Artif. Intell.*, 2010, pp. 111–120.

[16] F. Bonnin-Pascual and A. Ortiz, "A novel approach for defect detection on vessel structures using saliency-related features," *Ocean Eng.*, vol. 149, pp. 397–408, 2018.

[17] K. Yao, A. Ortiz, and F. Bonnin-Pascual, "A weakly-supervised semantic segmentation approach based on the centroid loss: Application to quality control and inspection," *IEEE Access*, vol. 9, pp. 69010–69026, 2021.

[18] F. Liu and M. Fang, "Semantic segmentation of underwater images based on improved deeplab," *J. Mar. Sci. Eng.*, vol. 8, no. 3, Mar. 2020, Art. no. 188.

[19] B. C. Kim, H. C. Kim, S. Han, and D. K. Park, "Inspection of underwater hull surface condition using the soft voting ensemble of the transfer-learned models," *Sensors*, vol. 22, no. 12, 2022, Art. no. 4392.

[20] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. 9th Int. Conf. Learn. Representations*, 2021.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[22] A. Kanadath, J. A. A. Jothi, and S. Urolagin, "Histopathology image segmentation using mobilenetv2 based U-net model," in *Proc. Int. Conf. Intell. Technol.*, 2021, pp. 1–8.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9351. Berlin, Germany: Springer, 2015, pp. 234–241.

[24] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Representations*, 2015, pp. 1–14.

[26] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[27] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 936–944.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[29] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6230–6239.

[30] M. Abadi et al., "Tensorflow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Oper. Syst. Des. Implementation*, 2016, pp. 265–283.

[31] M. Själander, M. Jahre, G. Tufte, and N. Reissmann, "EPIC: An energy-efficient, high-performance GPGPU computing research infrastructure," 2019, *arXiv:1912.05848*.

[32] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.

[33] M. O'Byrne, F. Schoefs, V. Pakrashi, and B. Ghosh, "An underwater lighting and turbidity image repository for analysing the performance of image-based non-destructive techniques," *Struct. Infrastructure Eng.*, vol. 14, no. 1, pp. 104–123, Jan. 2018.

[34] C. Li et al., "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

[35] *International Chamber of Shipping and the Baltic and International Maritime Council*, Industry Standard on In-Water Cleaning With Capture, vol. 1, London, U.K., Feb. 2021. [Online]. Available: https://www.ics-shipping.org/publication/industry-standard-on-in-water-cleaning-with-capture

[36] S. Hong and J. Kim, "Three-dimensional visual mapping of underwater ship hull surface using piecewise-planar slam," *Int. J. Control, Automat. Syst.*, vol. 18, no. 3, pp. 564–574, 2020.

Maryna Waszak (Member, IEEE) received the Ph.D. degree in electrical engineering from the École polytechnique fédérale de Lausanne, Lausanne, Switzerland, in 2016, and developed a new motion correction technique for magnetic resonance imaging.

She is a Research Scientist with SINTEF Digital Smart Data Group, Oslo, Norway. The expertise of the group focuses on the fusion of data of different modalities and from highly heterogeneous sources. Her current research work is tailored around applications in the processing industry in the context of digital twins.

Alexandre Cardaillac received the Bachelor of Information Technology from the School of Digital Innovation, Nantes, France, in 2019 and the M.Sc. degree in artificial intelligence with speech and multimodal interaction from Heriot-Watt University, Edinburgh, UK, in 2020. He is currently working toward the Ph.D. degree in engineering with the Department of Marine Technology, Norwegian University of Science and Technology, Trondheim, Norway.

Brian Elvesæter received the Cand.scient. degree in computer science from the University of Oslo, Oslo, Norway, in 2000.

He is a Senior Research Scientist with the Department for Sustainable Communication Technologies, SINTEF Digital. His current research interests include data management and data pipelines, knowledge graphs and semantic technologies, and AI and graph analytics.

Frode Rødølen received the Diploma from the Maritime Highschool, Bergen, Norway, in 1999.

He is the Founder and CEO of VUVI AS, Bergen, Norway. VUVI is a frontrunner in using ROV for surveys on ship hulls and is an approved supplier to the Norwegian Maritime Authorities, DNV, Lloyd's Register, Rina, CCS, and Bureau Veritas vessels. Since the commencement of operations, VUVI has executed more than 350 vessel inspections. He has more than 20 years of experience and has learned how to use and modify small inspection class ROV's to inspect ship hulls. He is passionate about lifelong learning, entrepreneurship, and social responsibility through nonprofit work, and he welcomes opportunities to explore, especially the ocean.

Martin Ludvigsen (Member, IEEE) was born in 1977. He received the Ph.D. degree in underwater technology from the Norges Teknisk-Naturvitenskapelige Universitet, Trondheim, Norway, in 2010.

Since 2014, he has been a Professor with the Department of Marine Technology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. He is a Cofounder and Manager for the Applied Underwater Laboratory (AUR-Lab) with NTNU. The AUR-Lab (https://www.ntnu.edu/web/aur-lab/aur-lab) facilitates research within both engineering disciplines and marine science by providing ROV, AUV, and USV operations. He has long experience at-sea both in arctic waters as well as in benthic environments associated with the Norwegian midocean ridge. He has been involved in research projects both in the deep sea, the upper water column, and arctic deploying robotic underwater vehicles. His research interests cover the field of underwater vehicles including perception and interpretation of cameras and sonar data together with autonomy. Adaptive mission planning for one or more vehicles for ocean column mapping has also been a focus point for his research group.