Hallucinating Hidden Obstacles for Unmanned Surface Vehicles Using a Compositional Model

Jon Muhovič[1], Gregor Koporec[1,2] and Janez Perš[1,*]

[1]Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, 1000 Ljubljana, Slovenia
[2]Gorenje, d.o.o., 3320 Velenje, Slovenia

Abstract

The water environment in which unmanned surface vehicles (USVs) navigate presents many unique challenges. One of these is the risk of encountering obstacles that are (partially) submerged and therefore poorly visible. Therefore, their extent cannot be determined directly from available above-water sensor data. On the other hand, it is well known that human skippers are able to safely navigate boats around obstacles even without underwater sensors and only with the help of their expertise. In this paper, we describe initial work on extending the USV obstacle detection to include such functionality using a compositional model. To learn to hallucinate the extent of obstacles with a minimum of learning effort, we exploit the nature of obstacles (people in kayaks, canoes, and on paddleboards) that are visible most of the time, but not always. We evaluate the impact of such hallucinations on USV safety and maneuverability, and suggest additional cases where such hallucinations can be used to improve USV safety.

Keywords

unmanned vehicles, USV, obstacle detection, compositional models

1. Introduction

Unmanned surface vehicles (USVs) are increasingly recognized as a valuable tool for a variety of applications, including military, environmental, and commercial purposes. These autonomous craft are capable of operating in difficult or hazardous environments, making them ideal for tasks that would be too risky for humans.

On the other hand, one of the envisioned benefits of USVs is the ability to gather data and perform tasks for extended periods of time without the need for human intervention. This would allow them to cover large areas and collect a large amount of data that can then be used for a variety of purposes. USVs equipped with sensors and cameras could be used, for example, to monitor and map the marine environment, track wildlife [1], or assess the health of coral reefs [2]. However, truly autonomous vehicles with no captain on board and no contact with remote operators must essentially duplicate the reasoning of a trained skipper in certain situations. One of those situations are (partially) submerged objects that cannot be detected by USV sensors located above the water, but whose presence could be easily inferred by a human operator.

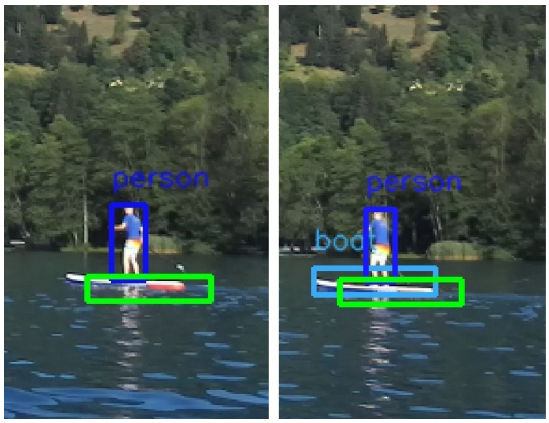

Figure 1: Left: detection of objects using Yolov7 [3], a person is detected (dark blue), but neither boat nor paddle board are detected. We hallucinate the boat (in green). Right: Same person, later, when the boat is actually detected by Yolov7 (light blue), comparing the actual detection versus the hallucination (green).

Our approach is best illustrated by observing Fig. 1. Based on the observation that people cannot walk or sit on water, we force the hallucination of a boat with every person that is detected on the surface of the water. The parameters of the hallucinated object are learned from person-boat compositions obtained by using a pre-trained object detector on a separate dataset and do not require annotation.

26th Computer Vision Winter Workshop, Robert Sablatnig and Florian Kleber (eds.), Krems, Lower Austria, Austria, Feb. 15-17, 2023
*Corresponding author.
†These authors contributed equally.
jon.muhovic@fe.uni-lj.si (J. Muhovič); gregor.koporec@gorenje.com (G. Koporec); janez.pers@fe.uni-lj.si (J. Perš)
https://lmi.fe.uni-lj.si/en/jon-muhovic/ (J. Muhovič); https://lmi.fe.uni-lj.si/en/janez-pers-2/ (J. Perš)
0000-0002-6039-6110 (J. Perš)

This paper is organized as follows: Following related work, we define the problem we want to use to demonstrate the capabilities of our method. We then introduce the basic concepts of compositional models and describe our use case and evaluation method. In the experimental part, we present our own dataset used in our experiments and its properties, followed by the evaluation setup focusing on the USV navigation. Finally, we discuss the results and further applications of the presented approach.

2. Related work

Recently, numerous papers have been published on the subject of USV sensors, obstacle detection and navigation.

The computer vision aspect of marine environment interpretation has been approached in several ways so far: Some authors have acquired datasets to facilitate domain transfer for Deep Learning and further investigate the specific problems in the maritime domain [4, 5, 6]. Several USV architectures with different sensors have been presented to solve problems such as poor lighting conditions and the need for absolute distance measurements [7, 8, 9]. In addition, authors have proposed deep learning methods specific to the maritime domain that either incorporate additional relevant modalities or address problems that arise in the maritime domain [10]. Numerous publications have also been presented that address automatic navigation and maritime collision avoidance compliance [11].

Han et al. [12] have presented a complete platform and framework for obstacle detection and avoidance, complete with multimodal sensors, obstacle detectors, and collision avoidance rules. They use SSD detector [13] to detect potential obstacles and track them using sensor fusion. Since real-time performance is usually desired, fast detectors such as SSD or YOLO [14] are usually preferred for USV applications.

Several datasets have also been published, some of which will be used as learning data for Deep Learning-based methods and others as benchmarks for existing methods. One such dataset, SMD, was proposed by Prasad et al. [4]. It contained 51 RGB and 30 NIR sequences and was primarily intended for monitoring. Since then, several more USV oriented datasets have been proposed, such as MODD [15], MaSTr1325 [5], and MODS [6].

In the past, obstacle detection was performed directly by estimating salient regions [16] or color segmentation [8]. Before the widespread use of Deep Learning, several approaches were also proposed that mainly focused on semantic segmentation followed by anomaly detection. These methods [15, 17] used prior information about the scene and refined it with color image information. With the advent of Deep Learning, the two branches of obstacle detection have been improved. On the one hand, researchers have adapted or retrained general object detectors for marine environments [18, 19, 10] using more precise classification information and custom datasets. However, such approaches only work for well-defined objects. Unknown structures, such as floating debris or piers, cannot usually be detected using such methods.

The other branch of obstacle detection is semantic segmentation. Several methods have adapted general segmentation methods to the marine environment [20, 7, 21]. Obstacle detection can be performed using such methods by determining regions that are partially or completely surrounded by water.

The method presented in this paper operates at a higher level of reasoning and aims to use assumptions that reasonably hold in water-bound environments. It relies on existing but imperfect methods for obstacle detection (in this paper we use Yolov7 [3]).This work contains two contributions:

- A method for improving the safety of USV and its environment by improving the estimation of free passage corridors in front of the USV, even with imperfect obstacle detectors.
- An evaluation method that evaluates increase of safety in that case

3. Problem definition

In situations where we cannot reliably observe fully or partially submerged obstacles using any of the sensors mounted above the water, we use knowledge of commonly occurring structures in marine environments to improve the safety of a USV.

In this paper, we present preliminary research results: We focused on the problem of *detecting boats or other floating objects in situations where a person was detected above the water surface, but the corresponding boat was not detected.* Such cases often occur when boats are of a similar color to the surrounding water, partially submerged due to maneuvering, or are otherwise poorly visible due to backlight or the distance between a smaller object and the camera. The work was performed using RGB images, because of the wide availability of pre-trained object detectors that perform reasonably well without the need for additional training.

Since we are dealing with coastal and continental water regions where smaller boats such as rowboats and paddle boats are usually found, consistent detection of such obstacles is necessary. Depending on the lighting conditions, size and color of the boats, detection with conventional detectors applied to color images is not always consistent. This inconsistency can be a hazard to

safe navigation, especially when maneuvering near other boats.

This problem has the following interesting properties:

- Solid physical foundation. People cannot walk or sit on the water. There must be some kind of highly buoyant device present to support their weight.
- No opportunity to *introduce* gross errors with false detections. False positives only restrict the possibilities for the USV to advance, and our experiments were designed to check for that effect.
- No manual annotations are needed, since we can obtain ground truth using the object detector (Yolov7) and therefore obtain plenty of data to train the higher-reasoning model.

The method will be later extended to a wider range of problems, which are discussed in Section 7.2, but represent edge cases and thus are subject to problems of data collection.

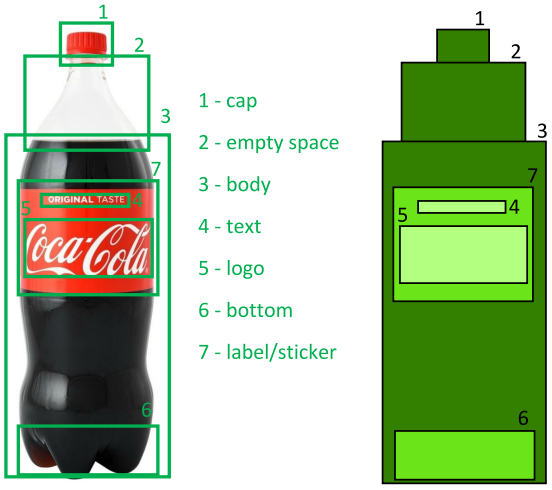

Figure 2: Concept of the compositional model – modelling of the Coke bottle. The composition is shown on the left, each part marked with a green rectangle. Names of the parts are shown in the middle. A compositional *hierarchical* model is shown on the right: darkest rectangles represent high level parts 1, 2, 3. Of those, part 3 is a composition itself, containing parts 6 and 7 (lighter). Part 7 is again a composition of parts 4 and 5 (lightest).

4. Our method

Our method is heavily influenced by the work of Koporec et al. [22], that uses hierarchical compositional models to detect objects' visible parts even when large parts of objects are occluded, and allows collection of expert knowledge from a small number of targeted human annotations. In our work we use a highly simplified implementation of Human-Centered Deep Compositional (HCDC) model [22].

4.1. Compositional models

In computer vision, *a composition* refers to the arrangement of visual elements in an image. These visual elements are called *parts* and can be low level primitives (e.g. edges, corners) or high-level objects themselves (e.g. cap, a label and recognizable shaped bottom on a bottle of soft drink), as shown in Fig. 2. Parts can be compositions themselves, yielding a hierarchical compositional model.

The compositional model, as shown in Fig. 2 is not particularly useful, as it is completely rigid. In practice, geometrical parameters of the parts are modelled as random vectors. In Figure 3 we show a hierarchical, compositional model of a 3-part Coke bottle under the assumption that the probability distribution of j-th part position (x_{ij}, y_{ij}) relative to the center (origin) of the j-th composition is Gaussian:

$$\begin{aligned} \mathbf{X}_{ij} &= \begin{bmatrix} x_{ij} & y_{ij} \end{bmatrix}^T \\ \mathbf{X}_{ij} &\sim \mathcal{N}(\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}) \end{aligned} \tag{1}$$

where $\mathbf{X}_{ij}$ is a two-dimensional random vector, generated by the Gaussian distribution $\mathcal{N}_2$ with mean vector $\boldsymbol{\mu}_{ij}$ and covariance matrix $\boldsymbol{\Sigma}_{ij}$. The parameters of the Gaussian distribution are obtained by learning on a sufficiently large set of train data, from which vectors $\mathbf{X}_{ij}$ are extracted.

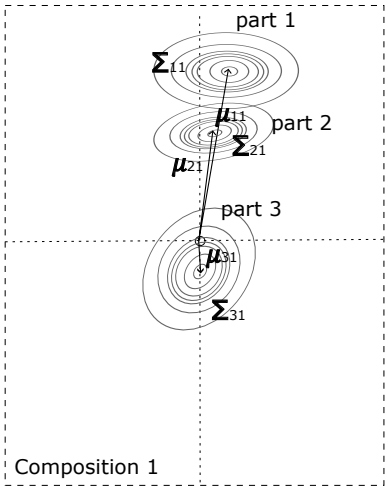

Figure 3: Illustration: three parts of a Coke bottle (parts 1, 2, and 3 from Fig. 2 could look something like this, if the learning samples would feature coke bottles tilted slightly to the right. Other parts are not shown. Ellipses show Gaussian distributions of part displacements (x_{ij}, y_{ij}) relative to the center of the composition (denoted as Composition 1).

Compositional models can be used in the following ways:

- 1. Robust, explainable detection of partially occluded objects, where the object (composition) is detected even if not all its parts are visible.
- 2. Explanation (hallucination) of the missing part. This is the functionality we use in the presented work.

4.2. Model of a person on a boat

Human-Centered Deep Compositional (HCDC) model [22] operates on parts that are itself deep detections (detections, obtained by convolutional neural network models, CNNs). This makes the model explainable, as the parts are already categorized into human-understandable categories.

We follow this example and use the detections provided by an obstacle detector pretrained on MS COCO [23]. We only retained the pertinent detection classes: *person*, *boat* and *surfboard*. Additionally, we treated the classes *boat* and *surfboard* as the same semantic entity (referred to as *boat* in the remainder of the text), since both of those classes almost always appear simultaneously with the class *person*. The compositional model that we use is shown in Fig. 4.

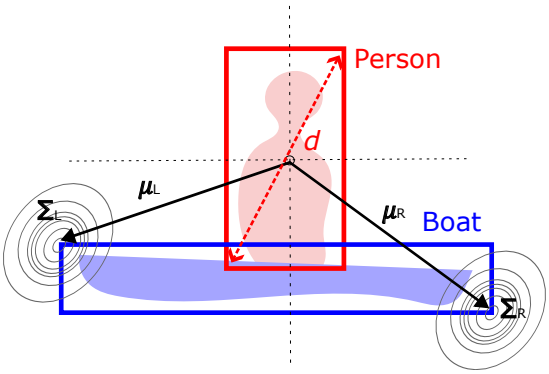

Figure 4: Model of a two-part composition we use in this research – a person on a boat. The centroid of the person detection is the origin of the composition coordinate system, and two corners of a bounding box represent the boat. The position of the corners is modelled using two Gaussian distributions. To adjust the model to different scales, we use the diagonal of the person's bounding box, d.

In our case, the Eq. (1) changes, since we have two separate Gaussian models for upper-left and bottom-right corners of the boat bounding box, and that for each of N scales.

$$
\begin{aligned}
d &= k\frac{d_{max}}{N} \\
\mathbf{X}_{Lk} &= \begin{bmatrix} x_{Lk} & y_{Lk} \end{bmatrix}^T \\
\mathbf{X}_{Rk} &= \begin{bmatrix} x_{Rk} & y_{Rk} \end{bmatrix}^T \\
\mathbf{X}_{Lk} &\sim \mathcal{N}_{Lk}(\boldsymbol{\mu}_{Lk}, \boldsymbol{\Sigma}_{Lk}) \\
\mathbf{X}_{Rk} &\sim \mathcal{N}_{Rk}(\boldsymbol{\mu}_{Lk}, \boldsymbol{\Sigma}_{Rk})
\end{aligned}
\tag{2}
$$

where subscripts L and R denote left-top or right-bottom point of the boat bounding box, respectively, and k denotes the scale index. Therefore, a total parameter set of our 2D model consists of $2N$ Gaussian means and $2N$ 2D Gaussian covariance matrices.

4.3. Training the compositional model

Our training does not require any manual annotations. Due to pretty good (but not perfect) performance of the chosen detector (Yolov7 detects about 95% boats and even higher percentage of persons) we use those cases where both the boat and the person on it were detected, to establish a model that can reasonably predict the position and size of a boat in absence of detections.

Although we assume Gaussian model for probability distributions $\mathcal{N}_{Lk}$ and $\mathcal{N}_{Rk}$, we estimate each separate distribution using expectation maximization (EM) algorithm with 2-component Gaussian Mixture Model (GMM) and retain the larger of the two components as either $\mathcal{N}_{Lk}$ or $\mathcal{N}_{Rk}$. Our preliminary testing has revealed that using 2-component GMM results in more accurate fitting of Gaussian model to the data, collecting the outliers in the significantly smaller component.

4.4. Hallucination

To hallucinate the most likely bounding box of the (undetected) boat, we examine the bounding box of the detected person, calculate its centroid and diagonal d, calculate the scale index k and look up the relevant Gaussian models $\mathcal{N}_{Lk}$ and $\mathcal{N}_{Rk}$ obtained during the training. The hallucinated bounding box points of a boat are determined at displacements x_{Lk} and y_{Lk} at which $\mathcal{N}_{Lk}$ and $\mathcal{N}_{Rk}$ have their maximum values. Note that x_{Lk} and y_{Lk} are relative to the person's centroid point.

5. USV safety-focused evaluation

To compare performance of object detectors, a generic approach by counting false positives and false negatives, with respect to some minimum intersection over union (IoU) value is often used. However, when evaluating the detectors in with actual application in mind, it is often the case, that not all errors are equally important or relevant. For example, USV benchmark [6] defines a

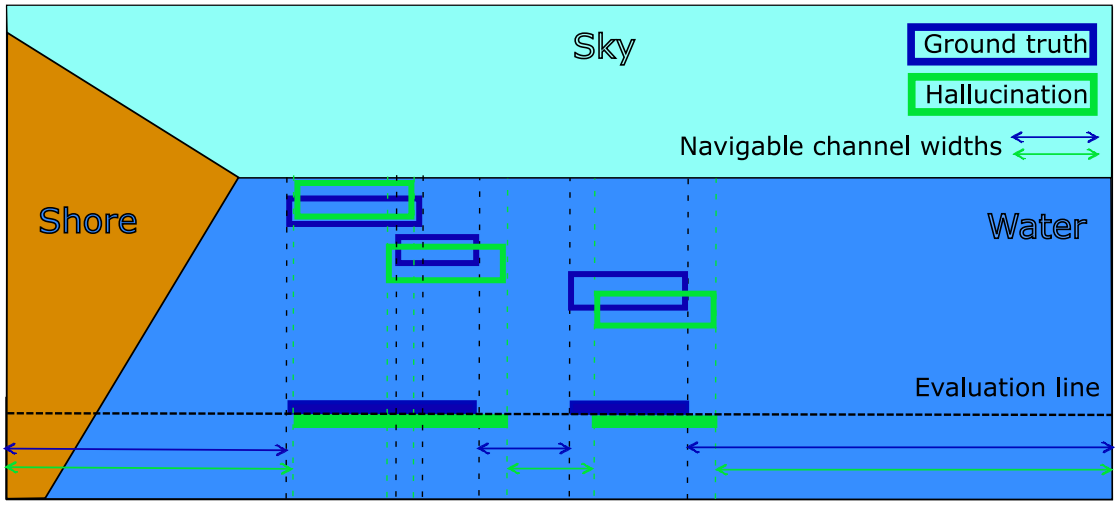

Figure 5: Illustration of the evaluation methodology. Note that the shore does not influence evaluation in any way, this is an intentional simplification. Blue and green bounding boxes represent ground truth detections and output of our method (hallucinations), respectively. All bounding boxes are *vertically projected* onto the horizontal (x) axis. All evaluation, including IoU is done in one dimension, along the horizontal axis. Arrows denote the widths of "navigable channels" after the projection of the bounding boxes onto the horizontal axis.

so-called *danger zone* to evaluate more relevant obstacles separately. The problem that we are addressing in this work is increasing safety of the USV navigation, in cases where actual boats are not detected. The challenge is, *how do we measure increase in safety?*

Note that a crucial safety issue here is that the USV may navigate in the areas that actually contain part of the boat. Fig. 5 shows the situation with multiple detections and corresponding hallucinations. The aim of the USV is to proceed in the forward direction, but it has to avoid obstacles. Therefore, it can proceed only through *navigable channels*, marked with arrows in Fig. 5. To ensure safety, navigable channels cannot contain any part of the boat at any distance, and the problem can be compressed to one-dimensional representation along the horizontal (x) axis. However, if the hallucinations are too wide, there may not be any navigable channel left in front of the boat.

Therefore, we define the following two metrics:

- One-dimensional IoU value (referred to as IoU-1D), calculated from the projections of actual (ground truth) bounding boxes and hallucinated bounding boxes, both projected downwards onto the horizontal axis (evaluation line in Fig. 5). This value should be as high as possible.
- One-dimensional coverage (referred to as Cov-1D) of the horizontal axis (evaluation line) with projection of both ground truth bounding boxes and hallucinated bounding boxes. If the coverage of hallucinations becomes too high, then USV may not have any possibility of advancing, and regardless of the increase of safety, this solution is not good. Coverage is obtained by dividing the width of the evaluation line in pixels with the sum of the pixels on the evaluation line, covered by projected bounding boxes.

This evaluation protocol does not assume or require complex obstacle avoidance maneuvers, and is not sensitive to vertical displacement of bounding boxes.

6. Experiments

We recorded several hours of video on the Ljubljanica river (sessions denoted LJU1, LJU2, and LJU3) in different weather conditions, on Lake Bled (denoted BLE1), and on the Adriatic Sea (near the coast, in several areas between Koper and Portorož), denoted ADR1. In each case, we hired human workers who served as obstacles in boats, kayaks, canoes and on paddleboards. The data contains about 10 obstacles in the near vicinity of the recording boat, captured in different configurations and from different angles relative to the position of the sun (so challenging backlit scenes were also captured). Videos were recorded at 10 frames per second using Stereolabs ZED 3D stereo camera[1], mounted between 1-1.5 meters above the water surface (different watercraft were used at different locations). In this experiment we only use the

[1] https://www.stereolabs.com

left RGB images, the right RGB image and depth were not used in any way.

6.1. Analysis of dataset contents

The training data was constructed by first obtaining predictions for all the relevant classes using Yolov7. The compositions were then constructed from cases where there was overlap between detections of class *person* and either of the classes *boat* or *surfboard*.

Analysis of the detections provide some insight into the problem of "invisible" boats and paddle boards, as shown in Table 1.

Session (dataset)	LJU1	LJU2	LJU3	BLE1	ADR1
person only (%)	0.04	0.05	0.05	0.03	0.05
person+boat (%)	0.96	0.95	0.95	0.97	0.95

Table 1
Percentages of *detected people without boats* vs detected people *with boats* among all detections for each of recording sessions. Note that the percentage of missing boat detections ranges from 3-5%. The videos contained negligible amount of people on the shore (physically plausible detections without boats).

6.2. Training

We decided to use session BLE1 for training of the Gaussian distributions $\mathcal{N}_{Lk}$ and $\mathcal{N}_{Rk}$, as it featured boats of varying shapes and sizes. The training time using precalculated Yolov7 detections was negligible.

6.3. Testing

Free from requirements for manual annotation, we were able to run the evaluation of our method on all images from our dataset, For evaluation, we used only the detections of people with corresponding boats. Boat detections, obtained via Yolov7, were considered ground truth, against which the hallucinations, obtained using our compositional model, were tested. Person detections without corresponding boats were not used, as these had no usable ground truth. Table 2 shows the results.

Analysing the results, we can see that there is good overlap between ground truth detections and hallucinations, with IoU-1D ranging from 0.465 to 0.605 for the same dataset on which the model was trained. Note that IoU-1D of 0.5 means that the middle half of bounding box projections overlap, while the 1/4 at each edge is non-overlapping.

Coverage of hallucinations is not as high as coverage of detections, and, most surprisingly, coverage of pure person detections (e.g. in absence of any detected boats) is not much lower than the coverage of hallucinations. We examined the reason behind this and found that the increase is not as high as expected due to obstacles which are further away and have disproportionately wide person detection bounding boxes, and due to differences in the set of boats used for training and testing (note the highest increase in Cov-1D from person detection to hallucination when the training set BLE1 was tested). Figure 6 shows an image where the result of our method is poor.

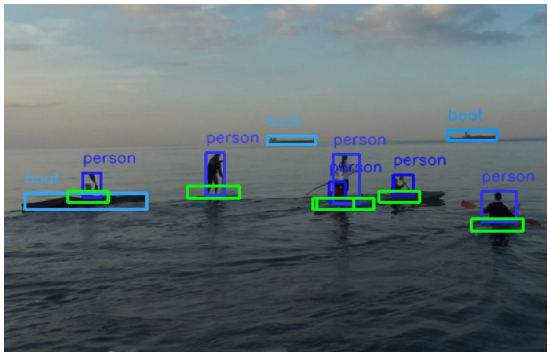

Figure 6: Image on which the proposed method does not significantly improve safety. Note the wide detections of persons and an uncharacteristically long canoe.

7. Discussion

This paper presents a preliminary research on use of hallucinations, provided by compositional models, in water-borne obstacle detection and avoidance. The experimental design in this work has been subject to many constraints, most notably the absence of proper ground truth annotations. These issues will be addressed in further work, towards a general framework to hallucinate obstacles that are not directly observed by the sensors.

Since using an obstacle detector precludes us from detecting unknown objects, combining their results with either semantic segmentation or another method of anomaly detection or a different sensor modality (such as LIDAR) might help in producing a more general hazard detection system that will perform hazard detection from multimodal cues.

7.1. Underwater sensors

A state of the art in experimental autonomous road vehicles relies heavily on multimodal sensor setup, with sensors like LIDAR and RADAR [24, 25], which bear no resemblance to human sensing. Therefore, an argument could be made that instead of hallucinating the obstacles and trying to emulate the skipper, one could detect the hidden obstacles using proper underwater sensor setup.

Session (dataset)	LJU1	LJU2	LJU3	BLE1	ADR1
IoU-1D	0.465	0.435	0.532	0.605	0.582
Ground truth Cov-1D	0.13	0.149	0.152	0.101	0.193
Hallucination Cov-1D	0.074	0.065	0.117	0.083	0.149
Person detection Cov-1D	0.067	0.054	0.127	0.062	0.139

Table 2
Evaluation results using the model trained on BLE1 session. IoU-1D is 1-dimensional IoU on bounding box projections onto the horizontal axis and Cov-1D is coverage of the horizontal axis with each type of bounding boxes. We included the projections of pure person detection bounding boxes as well for comparison.

In practice, this results in fragile setup due to water turbidity – USVs are expected to navigate safely even in water that is dirty or muddy.

Note also that a paddleboard, as shown in Fig. 1, is a very thin object at the boundary between air and water, which is not comparable to the situations encountered in autonomous driving (on the road), so it is unlikely that additional (underwater) sensors will reliably detect it. In fact, some watercraft may be completely submerged at times, as can be seen in Fig. 7 which shows a fast-moving athlete in a kayak.

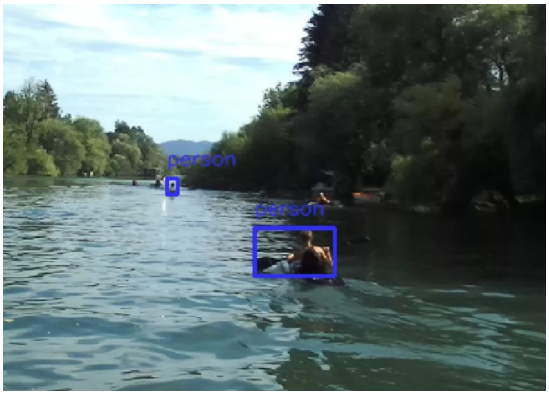

Figure 7: A submerged kayak that cannot possibly be reliably detected using visual sensors.

7.2. Other examples of invisible hazards

Missing detections of boats and paddleboards are immediately available in our waterborne datasets. However, there are other scenarios where such an approach would be useful, but for which there is currently insufficient data to train the models. The main reason for this is that these scenarios are to some extent hazardous to the USV and represent edge cases in USV deployment. In Figure 8, we present a common scenario that we have encountered several times, but for which we currently do not have enough data to properly test, let alone train. Plant debris is common in continental waters and usually safe to traverse. Often it covers the entire navigable area (e.g., leaves in the fall), so avoiding it at all times is not an option. However, debris may accumulate in shallow water areas (it may not be debris, but aquatic plants sticking out of the shallow water). So, if we encounter debris farther from shore, it is not a cause for concern as it is most likely floating. However, if it is found near land features (e.g., trees, mud), then it usually means that the area is dangerous, shallow, and not navigable. To detect this case, we might model the shallow, non-navigable area as a composition of debris and other land-based features.

As it can be seen in top right image in Fig. 8, it is sometimes difficult to determine whether the situation is a hazard or not. The labeling of such situations cannot be done by (untrained) labelers, but must be defined by experienced skippers working in cooperation with computer vision engineers. These compositions and their parameters must be defined by hand for a small number of available cases. The HCDC approach [22] has shown that this is indeed possible for common, well known food items. In this case, it will be used to insert *concentrated expert knowledge* into the compositional hazard detection model.

8. Acknowledgments

This work was financed by the Slovenian Research Agency (ARRS), research program [P2-0095], and research project [J2-2506].

References

[1] A. Dallolio, H. B. Bjerck, H. A. Urke, J. A. Alfredsen, A persistent sea-going platform for robotic fish telemetry using a wave-propelled usv: Technical solution and proof-of-concept, Frontiers in Marine Science 9 (2022). URL: https://www.frontiersin.org/articles/10.3389/fmars.2022.857623. doi:10.3389/fmars.2022.857623.

[2] G. T. Raber, S. R. Schill, Reef rover: A low-cost small autonomous unmanned surface vehicle (usv) for mapping and monitoring coral reefs, Drones 3 (2019). URL: https://www.mdpi.com/2504-446X/3/2/38. doi:10.3390/drones3020038.

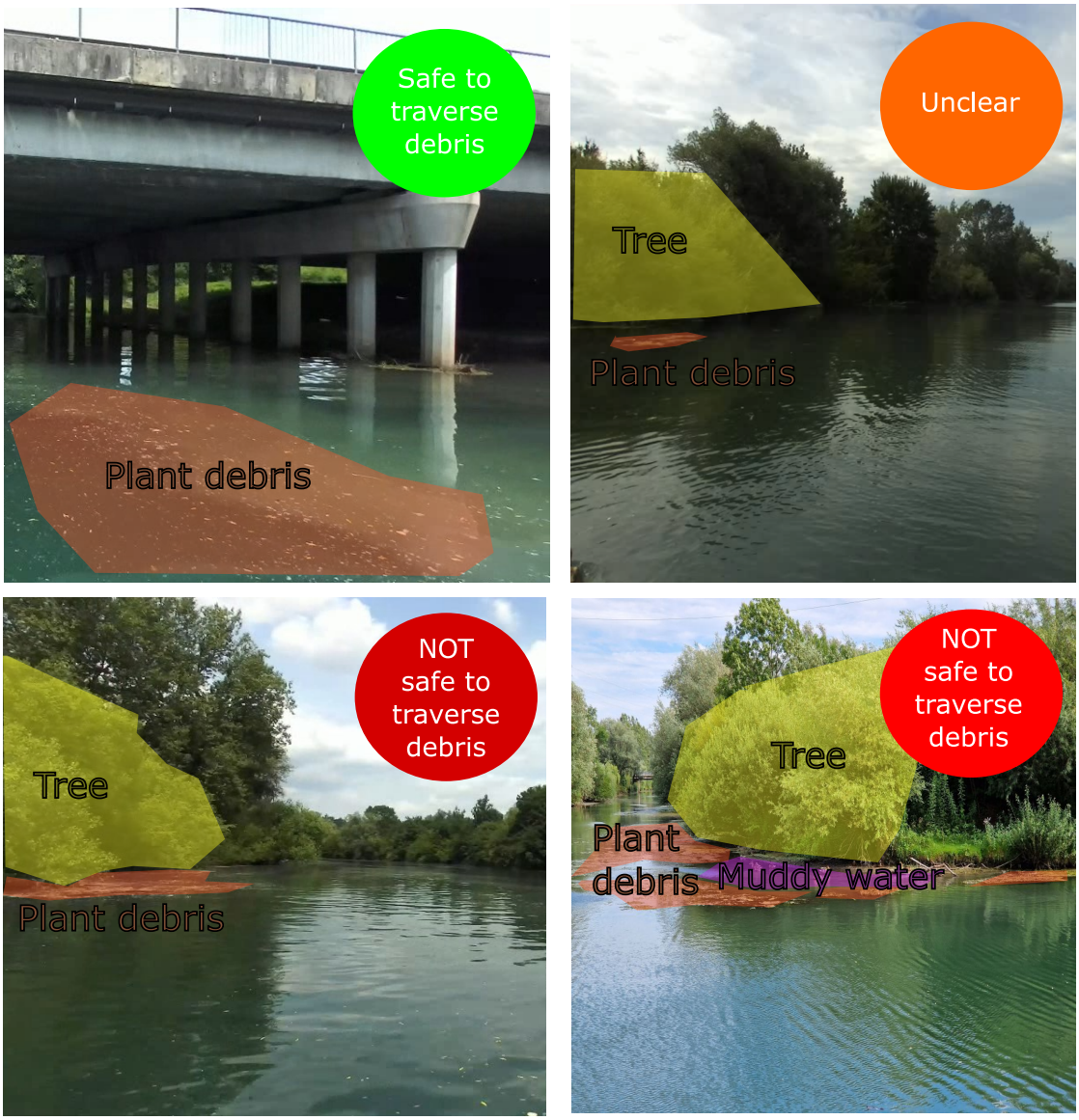

Figure 8: Example of an invisible danger - plant water debris. In all four images, plant debris can be seen in the image. Plant debris is usually mobile, buoyant, harmless, and can be run over by a boat (top left). However, if the plant debris is near the shore, it can accumulate on aquatic plants and signal dangerously shallow depth (top right and bottom left). The presence of other clues (muddy water) increases the likelihood that the water in the area of the debris is precariously shallow (bottom right).

[3] C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, 2022. URL: https://arxiv.org/abs/2207.02696. doi:10.48550/ARXIV.2207.02696.

[4] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, C. Quek, Video processing from electro-optical sensors for object detection and tracking in a maritime environment: a survey, IEEE Transactions on Intelligent Transportation Systems 18 (2017) 1993–2016.

[5] B. Bovcon, J. Muhovič, J. Perš, M. Kristan, The mastr1325 dataset for training deep usv obstacle detection models, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Sys-

tems (IROS), IEEE, 2019, pp. 3431–3438.
[6] B. Bovcon, J. Muhovič, D. Vranac, D. Mozetič, J. Perš, M. Kristan, Mods–a usv-oriented object detection and obstacle segmentation benchmark, IEEE Transactions on Intelligent Transportation Systems (2021).
[7] L. Steccanella, D. Bloisi, A. Castellini, A. Farinelli, Waterline and obstacle detection in images from low-cost autonomous boats for environmental monitoring, Robotics and Autonomous Systems 124 (2020) 103346.
[8] A. J. Sinisterra, M. R. Dhanak, K. Von Ellenrieder, Stereovision-based target tracking system for usv operations, Ocean Engineering 133 (2017) 197–214.
[9] Y. Cheng, M. Jiang, J. Zhu, Y. Liu, Are we ready for unmanned surface vehicles in inland waterways? the usvinland multisensor dataset and benchmark, IEEE Robotics and Automation Letters 6 (2021) 3964–3970.
[10] D. Nunes, J. Fortuna, B. Damas, R. Ventura, Real-time vision based obstacle detection in maritime environments, in: 2022 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), IEEE, 2022, pp. 243–248.
[11] Y. Kuwata, M. T. Wolf, D. Zarzhitsky, T. L. Huntsberger, Safe maritime autonomous navigation with colregs, using velocity obstacles, IEEE Journal of Oceanic Engineering 39 (2013) 110–119.
[12] J. Han, Y. Cho, J. Kim, J. Kim, N.-s. Son, S. Y. Kim, Autonomous collision detection and avoidance for aragon usv: Development and field tests, Journal of Field Robotics 37 (2020) 987–1002.
[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C.-Y. Fu, A. C. Berg, Ssd: Single shot multibox detector., in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), ECCV (1), volume 9905 of *Lecture Notes in Computer Science*, Springer, 2016, pp. 21–37.
[14] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, 2015. URL: http://arxiv.org/abs/1506.02640, cite arxiv:1506.02640.
[15] M. Kristan, V. S. Kenk, S. Kovačič, J. Perš, Fast image-based obstacle detection from unmanned surface vehicles, IEEE transactions on cybernetics 46 (2015) 641–654.
[16] H. Wang, Z. Wei, S. Wang, C. S. Ow, K. T. Ho, B. Feng, A vision-based obstacle detection system for unmanned surface vehicle, in: Robotics, Automation and Mechatronics (RAM), 2011 IEEE Conference on, IEEE, 2011, pp. 364–369.
[17] B. Bovcon, J. Perš, M. Kristan, et al., Stereo obstacle detection for unmanned surface vehicles by imu-assisted semantic segmentation, Robotics and Autonomous Systems 104 (2018) 1–13.
[18] J. Yang, Y. Li, Q. Zhang, Y. Ren, Surface vehicle detection and tracking with deep learning and appearance feature, in: 2019 5th International Conference on Control, Automation and Robotics (ICCAR), IEEE, 2019, pp. 276–280.
[19] S. Moosbauer, D. Konig, J. Jakel, M. Teutsch, A benchmark for deep learning based object detection in maritime environments, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 0–0.
[20] H. Kim, J. Koo, D. Kim, B. Park, Y. Jo, H. Myung, D. Lee, Vision-based real-time obstacle segmentation algorithm for autonomous surface vehicle, IEEE Access 7 (2019) 179420–179428.
[21] B. Bovcon, M. Kristan, A water-obstacle separation and refinement network for unmanned surface vehicles, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2020, pp. 9470–9476.
[22] G. Koporec, J. Perš, Human-centered deep compositional model for handling occlusions, 2022. 2nd revision in Pattern Recognition.
[23] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, P. Dollár, Microsoft coco: Common objects in context, 2014. URL: http://arxiv.org/abs/1405.0312, cite arxiv:1405.0312Comment: 1) updated annotation pipeline description and figures; 2) added new section describing datasets splits; 3) updated author list.
[24] J. Peršić, I. Marković, I. Petrović, Extrinsic 6dof calibration of a radar–lidar–camera system enhanced by radar cross section estimates evaluation, Robotics and Autonomous Systems 114 (2019) 217–230.
[25] C. Schöller, M. Schnettler, A. Krämmer, G. Hinz, M. Bakovic, M. Güzet, A. Knoll, Targetless rotational auto-calibration of radar and camera for intelligent transportation systems, in: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE, 2019, pp. 3934–3941.